Information Sharing Among Military Headquarters

The Effects on Decisionmaking

Walter L. Perry, James Moffat

AND NATIONAL SECURITY RESEARCH DIVISION

Information Sharing Among Military Headquarters

The Effects on Decisionmaking

Walter L. Perry
James Moffat

Prepared for the United Kingdom Ministry of Defense

RAND NATIONAL SECURITY RESEARCH DIVISION

A joint US/UK study team conducted the research described in this report. In the US, the research was carried out within RAND Europe and the International Security and Defense Policy Center of the RAND National Security Research Division, which conducts research for the US Department of Defense, allied foreign governments, the intelligence community, and foundations. In the UK, the Defence Science and Technology Laboratory (Dstl) directed the work and participated in the research effort. The RAND Corporation has been granted a licence from the Controller of Her Britannic Majesty's Stationery Office to publish the Crown Copyright material included in this report.

Library of Congress Cataloging-in-Publication Data

Perry, Walt L.
Information sharing among military headquarters : the effects on decisionmaking / Walter L. Perry, James Moffat.
p. cm.
"MG-226."
Includes bibliographical references.
ISBN 0-8330-3668-8 (pbk. : alk. paper)
1. Command and control systems—United States. 2. United States—Armed Forces—Communication systems. 3. Military art and science—United States—Decision making. 4. United States—Armed Forces—Headquarters. I. Moffat, James, 1948– II. Title.

UB212.P49 2004
355.3'3041—dc22

2004018584

The RAND Corporation is a nonprofit research organization providing objective analysis and effective solutions that address the challenges facing the public and private sectors around the world. RAND's publications do not necessarily reflect the opinions of its research clients and sponsors.

Published 2004 by the RAND Corporation
1776 Main Street, P.O. Box 2138, Santa Monica, CA 90407-2138
1200 South Hayes Street, Arlington, VA 22202-5050
201 North Craig Street, Suite 202, Pittsburgh, PA 15213-1516
RAND URL: http://www.rand.org/
To order RAND documents or to obtain additional information, contact
Distribution Services: Telephone: (310) 451-7002;
Fax: (310) 451-6915; Email: order@rand.org

Preface

New concepts such as network-centric operations and distributed and decentralised command and control have been suggested as technologically enabled replacements for platform-centric operations and for centralised command and control in military operations. But as attractive as these innovations may seem, they must be tested before adoption. This report assesses the effects of collaboration across alternative information network structures in carrying out a time-critical task, identifies the benefits and costs of local collaboration, and looks at how 'information overload' affects a system.

A joint US/UK study team conducted the research described in this report. In the United States, the research was carried out within RAND Europe and the International Security and Defense Policy Center of the RAND National Security Research Division, which conducts research for the US Department of Defense, allied foreign governments, the intelligence community, and foundations. In the United Kingdom, the Defence Science and Technology Laboratory (Dstl) directed the work and participated in the research effort. Dstl is the centre of scientific excellence for the Ministry of Defence, with a mission to ensure that the UK armed forces and government are supported with in-house scientific advice. RAND has been granted a licence from the Controller of Her Britannic Majesty's Stationery Office to publish the Crown Copyright material included in this report.

This report will be of interest to military planners, operators, and personnel charged with assessing the effects of alternative infor-

mation network structures, processing facilities, and dissemination procedures. Planners contemplating the use of network-centric processes to achieve military objectives can use the methods described in the report to evaluate alternative structures and processes. Information technologists can assess the contribution of each alternative to the decisionmaker's knowledge prior to taking a decision. The ultimate goal is to develop tools that will allow operators to quickly evaluate plans for their level of situational awareness.

For more information on the RAND International Security and Defense Policy Center, contact the director, James Dobbins. He can be reached by email at James_Dobbins@rand.org; by phone at 310-393-0411, extension 5134; or by mail at RAND Corporation, 1200 South Hayes Street, Arlington, VA, 22202-5050. More information about RAND is available at www.rand.org.

Contents

Figures

Tables

Summary

New information technologies introduced into military operations provide the impetus to explore alternative operating procedures and command structures. New concepts such as network-centric operations and distributed and decentralised command and control have been suggested as technologically enabled replacements for platform-centric operations and for centralised command and control. As attractive as these innovations seem, it is important that military planners responsibly test these concepts before their adoption. To do this, models, simulations, exercises, and experiments are necessary to allow proper scientific analysis based on the development of both theory and experiment.

The primary objective of this work is to propose a theoretical method to assess the effects of information gathering and collaboration across an information network on a group of local decision-making elements (parts of, or a complete, headquarters). The effect is measured in terms of the reduction in uncertainty about the information elements deemed critical to the decisions to be taken.

Our approach brings together two sets of ideas, which have been developed thus far from two rather different perspectives. The first of these sets is the Rapid Planning Process, developed as part of a project on command and control in operational analysis models within the UK Ministry of Defence Corporate Research Programme. It is a construct for representation of the decisionmaking of military commanders working within stressful and fast-changing circumstances. The second set of ideas comes from the work on modelling the effects

of network-centric warfare, carried out recently by the RAND Corporation for the US Navy. We assess the effects of collaboration across alternative information network structures in prosecuting a time-critical task using a spreadsheet model. We quantify the benefits and costs of local collaboration using a relationship based on *information entropy* as a measure of local network knowledge. We also examine the effects of complexity and information overload caused by such collaboration.

Decisions in a Network

New technologies are enabling militaries to leverage information superiority by integrating improved command and control capabilities with weapon systems and forces through a network-centric information environment. The result is a significant improvement in awareness, shared awareness, and collaboration. These improvements in turn affect the quality of the decisionmaking process and the decision itself, which ultimately lead to actions that change the battlespace.

In this report, we focus on the quality of the decisions, or the planned outcome, rather than on whether or not the desired effect is eventually achieved.

We note that decisions are made based on the information available from three sources: information that is resident at the decision node; information from collection assets and information processing facilities elsewhere in the network; and information from other local decisionmakers with whom the decision nodes are connected and with whom they share information.

Rapid Planning Process

In most cases, decisionmakers must make decisions without full understanding of the values of the critical information elements needed to support the decisions. The decision taken depends on the current values of the critical information elements, which are dependent on the scenario. This dependency is modelled using the Rapid

Planning Process. The critical information elements map out the commander's conceptual space. In the basic formulation of the Rapid Planning Process, a dynamic linear model is used to represent the decisionmaker's understanding of the values of these factors over time. This understanding is then compared with one or more of the fixed patterns within the commander's conceptual space, leading to a decision.

A probabilistic information entropy model is used to represent the uncertainty associated with the critical information elements needed for the decision. Ideally, through the Rapid Planning Process, additional information from collection assets or from collaborating elements in the network serves to reduce uncertainty and therefore increase knowledge.

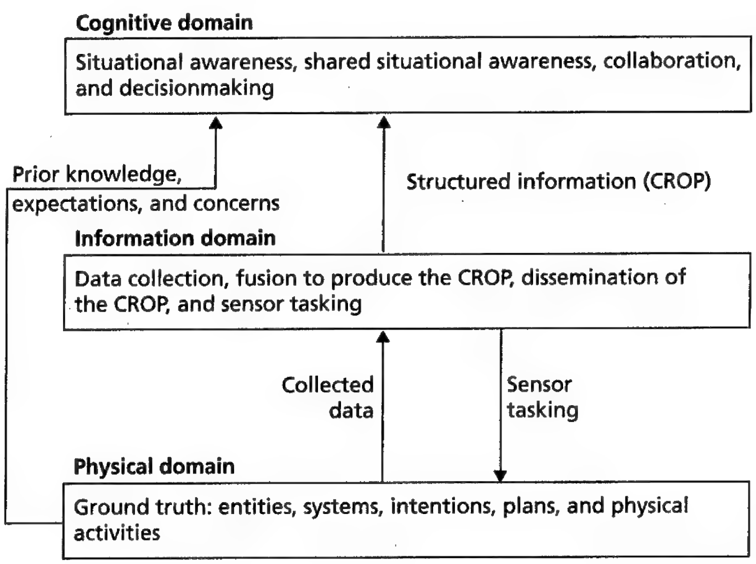
Knowledge

We are principally concerned with the information and cognitive domains, as depicted in Figure S.1. The domains of the information superiority reference model divide the command and control cycle into relatively distinct segments for ease of analysis. Their description includes the entities resident in the domain, the procedures performed and the products produced there, and the relationships among the domains.

Information derived from sensors or other information gathering resides in the information domain. This information is transformed into awareness and knowledge in the cognitive domain and forms the basis of decisionmaking. Our metrics quantify this process through the use of information entropy and knowledge measures.

Information sharing among nodes ideally tends to lower information entropy (and hence increase knowledge) partly because of the buildup of correlations among the critical information elements. That is, information can be gained about one critical information element (e.g., missile type) from another (e.g., missile speed). Such cross coupling is a key aspect for consideration, and we use conditional entropy to capture these relationships.

Figure S.1
The Information Superiority Reference Model



RAND *MG226-S.1*

Knowledge derived from entropy is a quantity that reflects the degree to which the local decisionmaker understands the values of the information elements. It is represented as a number between 0 and 1, with the former representing 'no understanding' and the latter representing 'perfect understanding'. From this knowledge, decisionmakers can assess whether or not they are in their 'comfort zone'—that is, whether the values of the key information elements support the decision they wish to take (such as one to launch the next attack mission).

Effects of Collaboration

Networks provide an opportunity for participating entities to share information as part of a collaborative process.[1] Here we focus on the synergistic effects of collaboration that improve the quantity (the completeness of our information) and the quality (its precision and accuracy) of the information needed to take decisions. We model the network as the combination of clusters of entities and represent each entity by a node. A cluster consisting of a single node is taken to be the degenerate case. Each such cluster consists of a set of entities, which have full shared awareness. *Full shared awareness* means that all entities in the cluster agree on the set of information elements and their values at any given time.

Estimators

Through observations of the battlespace, sensors and other information sources generate estimates for the information elements deemed critical to the decision. The uncertainty associated with the information elements is expressed in terms of probability distributions, the means of which are estimates of the ground-truth values. Because the mean of a probability distribution is a parameter of the distribution, we turn to parameter estimation theory to assess the quality of the information available to the decisionmaker and examine how the quality of the estimates contributes to knowledge.

- **Bias:** Bias in an estimate is error introduced by systematic distortions. An unbiased estimator is one for which its statistical expectation is the true value of the estimated parameter. That is, the expected value of the estimate of the parameter, $\hat{\mu}$, is the true value of the parameter, μ. The bias in the estimate is therefore the degree to which this is not true.
- **Precision:** The variation in estimates of the critical information elements can occur in a purely random way. Random errors

[1] *Collaboration* in this context is taken to be a process in which operational entities actively share information while working together towards a common goal.

affect the precision of the estimates reported because they increase the variance of the distribution of the estimated information element. In general, precision is defined to be the degree to which estimates of the critical information element or elements are close together.[2] Bias and precision, therefore, are independent—that is, biased estimates may or may not be precise.

Precision and Entropy

The amount of information available in a probability density is measured in terms of information entropy, denoted $H(x)$. Information entropy is always a function of the distribution variance, and therefore we use it as the basis for developing a knowledge function. For example, the bivariate normal distribution is $H(x,y) = \log|\boldsymbol{\Sigma}|$, where $\boldsymbol{\Sigma}$ is the covariance matrix. From this, we create a precision-based knowledge function as[3]

$$\begin{aligned} K(x,y) &= 1 - e^{-\left[\log|\boldsymbol{\Sigma}|_{\max} - H(x,y)\right]} \\ &= 1 - \frac{|\boldsymbol{\Sigma}|}{|\boldsymbol{\Sigma}|_{\max}} \end{aligned},$$

where $|\boldsymbol{\Sigma}|_{\max}$ is the determinant of the covariance matrix that produces the maximum uncertainty. Based on precision alone, $K(x,y)$ reflects the level of understanding within a cluster of decisionmakers.

For the simple case of two collaborating decisionmakers (i.e., two nodes of the network forming a cluster) who share two pieces of information with a multivariate normal distribution, the change in knowledge is given by

[2] This is a commonly accepted definition. Ayyub and McCuen (1997, p. 191) define precision as 'the ability of an estimator to provide repeated estimates that are very close together'. A similar definition can be found in Pecht (1995).

[3] Actually, the exact entropy value for the bivariate normal case is $H(x,y) = \log\left|(2\pi e)^2 \Sigma\right|$. However, because we are concerned about the relative entropy, we use the simpler version, which we refer to as 'relative entropy'.

$$\Delta K = \frac{\rho_{1,2}^2 \sigma_1^2 \sigma_2^2}{\sigma_{1,max}^2 \sigma_{2,max}^2},$$

where $\rho_{1,2}$ is the correlation coefficient, σ_1^2, σ_2^2 are the variances, and $\sigma_{1,max}^2, \sigma_{2,max}^2$ are the maximum or bounding values on the variance for the two pieces of information.

Accuracy

Accuracy is the degree to which the estimates of the critical information elements are close to ground truth. The concept of accuracy comprises both precision and bias. In general, if a is an information element whose value x is unknown with probability distribution $f(x)$ and mean μ representing ground truth, then the bias associated with the estimate of the mean is $b = |E(\hat{\mu}) - \mu|$, where $\hat{\mu}$ is the estimate of the mean. Because accuracy consists of both bias and precision, we therefore need a metric that combines both. One such metric is the mean square error (MSE), $E[(\hat{\mu} - \mu)^2] = b^2 + \sigma^2$, where σ^2 is the variance of $\hat{\mu}$. The MSE is an extremely useful metric because it includes both accuracy in the total and precision as a component. In estimating ground truth, the bias accounts for nonrandom errors and the precision accounts for random errors.

We illustrate by continuing with the bivariate normal case. We assume that Bayesian updating is used to refine the location estimate based on the arriving reports. Bayesian updating is not always unbiased, and therefore we introduce systemic error. In this case, the bias is the Euclidean distance between the Bayesian estimate and the ground-truth value:

$$b = \sqrt{\left(\hat{\mu}_x - \mu_x\right)^2 + \left(\hat{\mu}_y - \mu_y\right)^2}.$$

By analogy with the MSE, the accuracy of the estimate is defined as $D(x, y) = b^2 + |\hat{\Sigma}|$.

The Effects of Bias, Precision, and Accuracy on Knowledge

We now account for bias, precision, and hence accuracy in the knowledge function by replacing the distribution variance with the MSE, or the accuracy measure $D(x,y)$ in the knowledge function. Therefore, for the multivariate normal case, we get a modified knowledge function of the form:[4]

$$K_M(\mathbf{x}) = 1 - \frac{b^2 + |\Sigma|}{\left(b^2 + |\Sigma|\right)_{\max}}.$$

The 'maximum mean square error' is a combination of the maximum bias and the maximum precision and represents the maximum in inaccuracy. Because bias and precision are independent, the maximum occurs when both are maximised, or $(b^2 + |\Sigma|)_{\max} = b^2_{\max} + |\Sigma|_{\max}$. Like the variance, a suitable upper bound for bias can be found by searching for the largest possible measurement error the sensors or sources might produce.

Completeness

In addition to precision and accuracy, collaboration also affects the completeness of the critical information elements available within a cluster. For the entire network, we assume there are a maximum of N critical information elements. For a given cluster, the total number required is $C \leq N$. However, at a given time, t, only $n \leq C$ might be available. If waiting for additional reports is not possible, a decisionmaker would be required to take a decision without benefit of complete information. Depending on his experience and other contextual information, the decisionmaker may be able to infer some likely less reliable value for the missing information. For now, we assume that if the value of an information element is missing, the value of completeness at cluster i is

[4] The subscript M denotes knowledge derived from the MSE.

$$X_{i,t}(n) = \left[\frac{n}{C}\right]^{\xi},$$

where ξ is a 'shaping' factor. For values of $\xi < 1$, the curve is concaved downwards; for $\xi > 1$, it is concaved upwards; and for $\xi = 1$, it is a straight line. The selection of the appropriate value depends on the consequences associated with being forced to take a decision with incomplete information as well as the commander's attitude to risk.

Information Freshness

A final consideration when assessing uncertainty is that of freshness. The information arriving at a decision node consists of reports concerning one or more of the critical information elements necessary to take a decision. Both precision and accuracy depend on the joint probability density function that reflects the uncertainty in our knowledge of the ground-truth fixed pattern at a decision node. These reports are used to update the joint probability distribution of the information elements and hence the probability of correctness of each of the fixed patterns in the local decisionmaker's conceptual space.

We have selected Bayesian updating as the method for combining reports from various sources and sensors. All things being equal, we desire to give more weight to more recent reports, which requires that we reevaluate all available, valid reports at the time a decision is to be taken. A time-lapse estimate, $0 \le \Phi \le 1$, is used to determine the rate of information decay so that old information is given less weight than current information.

Measuring the Overall Effect of Cluster Collaboration

Finally, we combine the currency-adjusted precision and accuracy knowledge function with completeness to arrive at a single metric to assess the effects of collaboration across the cluster. The ideal case is when we have full completeness, i.e., $X_t(n) = X_t(C) = 1$, and the knowledge shared across the cluster is fully accurate, $K_M(x) = 1$. Unfortunately, this ideal is seldom, if ever, achieved. Consequently,

we require a construct that gauges the degree to which accuracy, as calculated here, and completeness contribute to knowledge.

In general, when $X_t(n)$ is small, the knowledge function should also be small. One way to reflect this behaviour is to replace the MSE in the entropy calculation with

$$\frac{b^2+\sigma^2}{X_t(n)}.$$

This equation has the desirable property that, when $X_t(n)\to 1.0$, the ratio is just the MSE, and when $X_t(n)\to 0$, it increases without bound. Because n is discrete, we can select $n=1$ to be the worse case, with $X_t(1)=C^{-\xi}$. Consequently, the upper bound on the resultant entropy calculation is

$$\frac{b_{\max}^2+\sigma_{\max}^2}{C^{-\xi}}=C^{\xi}\left(b_{\max}^2+\sigma_{\max}^2\right).$$

If $C=1$, there is no effect on the current entropy calculation or on the maximum entropy. If we let $K_\kappa(x)$ be the knowledge within the cluster based on accuracy and completeness, with the maximum variance replaced with $C^{\xi}(b_{\max}^2+\sigma_{\max}^2)$, we get

$$K_\kappa(x)=1-\frac{b^2+\sigma^2}{n^{\xi}\left(b_{\max}^2+\sigma_{\max}^2\right)}$$

for the univariate normal case.[5]

Up to this point, we have captured the effects of collaboration among decision nodes within a cluster on knowledge. The measured effects of information sharing through collaboration are accuracy and completeness. For the most part, these effects are dynamical, because they vary with the quality and quantity of reports received and processed over time. Missing from this analysis so far is an assessment of

[5] The κ subscript in this case refers to knowledge based on both the MSE and completeness.

the systemic effects of the network structure—that is, the effects that are more static. Next, we take up such measures of network complexity and combine them with the collaborative effects to arrive at a single measure of network performance and its effect on decision-making.

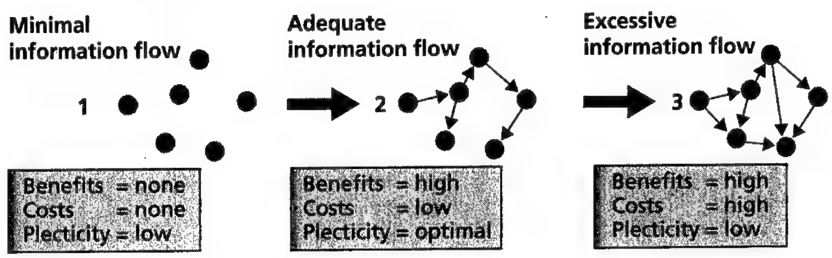
Effects of Structural Complexity

All networks exhibit complexity to a greater or lesser degree. Military command and control systems operating in a network-centric environment also exhibit complex behaviour. The challenge is understanding exactly what the complexity is, what its effects are, and how to quantify these effects. We note that there are both good and bad effects of complexity. Unfortunately, the term 'complexity' has a negative connotation; therefore, we have adopted Murray Gell-Mann's more neutral term, 'plecticity'.

In this context, *plecticity* refers to the ability of a connected set of actors to act synergistically via the connectivity between them. This measure is intended to take into account the fact that there may be constraints, due to technical or procedural limitations, on how nodes can constructively connect to other nodes; that is, a node's connectivity can add costs as well as benefits to the cluster. A measure of plecticity should account for the value of the cluster's ability to glean information from throughout the network to fulfil its particular functions, include a means for measuring the value of information redundancy, and reflect a cost to network effectiveness if nodes are overwhelmed.

For networks with inadequate clustering, as with excessive clustering—flows 1 and 3, respectively, in Figure S.2—we would expect low plecticity scores. The goal is to configure the information flow over a network with established link connectivity so as to maximise plecticity as measured in the terms discussed above and as illustrated by flow 2 in the figure.

Figure S.2
Overall Network Plecticity



RAND *MG226-S.2*

Accessing Information

The metric developed for completeness earlier is simply a ratio of counts: available required information elements to total required information elements. No attempt is made to assess the degree to which we can really *expect* to receive the information element, i.e., the degree to which the network allows the cluster to access information in the network. A metric that does so is the ratio of the aggregate expected degree of critical information access to the total number of required information elements. Such a metric accounts for the uncertainties associated with retrieving needed information.

We thus replace the binary accounting for information elements, with a connectivity score based on a distance function that recognises the cost imposed by the path the information must take through the network to arrive at the node requiring it.

For any information element, a_l, we are interested in the shortest path from source node to destination node, $d_l \geq 1$, however calculated. The restriction that the path distances always exceed 1.0 accounts for the fact that, for connectivity to exist at all, at least one link must exist between source and destination. The case in which no links exist implies an infinitely long path resulting in 0 connectivity. The quantity, d_l, represents the expense incurred by moving information element a_l from source to destination. The associated connectivity value is calculated as

$$k_l = \frac{1}{d_l^{\omega l}},$$

where $\omega_l \geq 1$ is the rate at which k_l varies with changing values of the distance function.

The strength of the connectivity among all the nodes in such a path must take into account the vulnerability of path elements (links and nodes) to attack or failure. We can do this using the connectivity score described above by examining its value as we remove each node—link or both—one at a time from a given path. For simplicity, we consider only the loss of nodes. We create a depletion vector, $\mathbf{L}_l$, whose elements consist of the connectivity values for information element a_l, with each of the path nodes removed in turn. The vector $\mathbf{L}_l$ then represents the vulnerability of the path and, as such, expresses the degree of uncertainty associated with retrieving information element a_l from network sources. The adjusted connectivity for information element a_l from network sources to a single destination is calculated to be

$$k_l^* = k_l \left(1 - \frac{\|\mathbf{L}_l\|}{\mathbf{L}_l}\right)^{\frac{1}{\rho}},$$

where $|\mathbf{L}_l|$ is the cardinality of the vector $\mathbf{L}_l$ and ρ is the edge expansion parameter of the network, which measures the overall robustness and reliability of the network. The resulting formula for accessibility, $X(k)$, is

$$X(k) = \begin{cases} \left(\dfrac{k}{C}\right)^{\xi} & C \neq 0 \\ 1 & \text{otherwise} \end{cases},$$

where $k = \sum_{l=1}^{C} k_l^*$ and C is, as before, the total number of information elements critical to the cluster.

Benefits of Network Redundancy

Network redundancy focuses on the reliability of the network; its ability to deliver information in the face of node loss; system outages; inefficient operating procedures; or some combination of all these elements. At the same time, a network can deliver excessive information, thus causing delays because of the time and resources required to process all of it. Consequently, network redundancy can be both a cost and a benefit of the network information flow.

Needed information can be provided to a cluster from multiple sources. If the value of the information will change over time, we can expect multiple reports from each source. These multiple reports require combining in some way as previously discussed under collaboration. Whatever method is used, the degree to which the reports contribute to estimates close to ground truth and to a narrowing of the distribution variance, a benefit will accrue to the cluster because of redundancy. Recall that the total number of required information elements across the whole network is N; the number critical to a cluster is C, where $C \le N$; and the number of these available within the cluster is n, where $n \le C$. If we let the vector $\boldsymbol{\Theta} = [\Theta_1, \Theta_2, \cdots, \Theta_C]^T$ represent the aggregate value of reports received for each required information element $(a_1, a_2, \cdots, a_C)$ from $\mathbf{P} = [p_1, p_2, \cdots, p_C]^T$ sources, then we can construct a suitable normalised aggregate metric, $R(\boldsymbol{\Theta})$, as

$$R(\boldsymbol{\Theta}) = 1 - \frac{1}{n} \sum_{i=1}^{C} \gamma_i e^{-\delta_i(\Theta_i - 1)},$$

where $\gamma_i = 1$ if $p_i \ge 1$ and 0 otherwise. We let $r_i(\Theta_i)$ be the benefit accruing from obtaining reports on the value of information element a_i from p_i sources where $\Theta_i = \sum_{j=1}^{p_i} \theta_{i,j}$, and $\theta_{i,j} \in [1, \infty)$ measures the assessed reliability of the report on information element a_i from source s_j. The parameter, δ_i, reflects the relative importance of the information element, a_i.

The combined benefit of information redundancy information to the cluster, based on the conditional dependency between accessibility and redundancy, is

$$B[R(\Theta)\,|\,X(k)]=\frac{(\beta-1)[\kappa X(k)+\beta R(\Theta)]}{(\beta-\kappa)[\beta-X(k)]},$$

where $\beta>1$ is a constant that ensures a nonzero denominator and $\kappa\geq 0$ is another constant that ensures that the combined metric is bounded between 0 and 1.

Costs of Information Overload

At the same time, a network can deliver excessive information. The more sources of required information and the more frequent the reporting, the longer it takes for the cluster to get a coherent view of the situation. That is, it takes time to process information, which may or may not contribute to improving the quality of the estimates. This excess is referred to as 'information overload'. In addition, some of the sources may provide disconfirming evidence. The value of the disconfirming evidence can be good or bad, depending on the degree to which it reflects ground truth. Disconfirming evidence requires time to evaluate and therefore may increase uncertainty and decrease the quality of the estimates. Finally, it is also possible that raw data may be processed before being sent, thus arriving at the cluster as time-stamped information with the time at which the processing ended. This possibility introduces an artificial latency that contributes to uncertainty.

The supply of unneeded information to a cluster has an immediate negative impact, because it must be processed or, at a minimum, interferes with the receipt of needed information. However, as more unneeded information is supplied, its impact is reduced. Thus, a good function to model this behaviour is the exponential $U(m)=1-e^{-\nu m}$, where m is the number of sources of unneeded information and ν is a scaling parameter.

The costs of information overload associated with needed information within a cluster are generally minimal for low levels of redundancy. Indeed, at these levels, the benefits far outweigh the costs, as discussed earlier. However, at some point, costs rise sharply so that the marginal cost of an additional source of information is

greater than the previous source. At some further point, this cost then levels off so that the marginal costs are minimal. This behaviour is best described using a logistics response function for each information element shared within the cluster. For simplicity, we express the combined costs of oversupply of needed information as a simple sum,

$$G(\mathrm{P}) = \frac{1}{n}\sum_{i=1}^{C} \delta_i \frac{e^{-(\chi_i+\varphi_i p_i)}}{1+e^{-(\chi_i+\varphi_i p_i)}},$$

where χ_i and φ_i are shaping parameters.

In considering the overall costs for the cluster, a balance is struck between costs of needed and unneeded information. We use a simple weighted linear sum of the two components of information overload, or $O[U(m),\mathrm{G}(\mathrm{P})] = \alpha U(m) + (1-\alpha)G(\mathrm{P})$, where $0 \leq \alpha \leq 1$, as a relative weight parameter.

Redundancy-Based Plecticity

The next step is to combine the costs and benefits of plecticity for a cluster associated with the mission at hand. For each cluster in the network, the measure of network plecticity, $C(B,O)$, is calculated as follows:

$$C(B,O) = B\left[R(\Theta)\middle|X(k)\right]\left[1-O\left[U(m),G(\mathrm{P})\right]\right].$$

Network Performance

The last step is to combine the redundancy-based plecticity with the benefits of collaboration across all the clusters of the network. Our collaboration metric quantifies the effects of information sharing across a cluster on information completeness and accuracy, whereas plecticity measures the positive and negative effects of redundant information and the degree of information access. The former assesses the dynamic nature of the operation conducted on the network; the

latter measures the effects of the underlying network structure and is therefore systemic. All the dependencies among the several components of collaboration and plecticity are not generally well understood. However, we know that high-quality performance requires good cluster knowledge and the means to share it and that scores in either category are penalised by deficiencies in the other. Therefore, the measure of total network performance is taken to be

$$\Omega(\Pi, \mathrm{K}_N) = \sum_{i=1}^{L} \left[C_i(B,O) K_{i,\kappa} \right]^{\omega_i},$$

where $\sum_{i=1}^{L} \omega_i = 1$ and L is the number of clusters.

For values of $\Omega(\Pi, \mathrm{K}_N)$ close to 1.0, the network is performing well by producing the information required to take decisions within each of the clusters when required. However, this is not the whole story. The next step is to assess how well the combat mission is accomplished. As important as good decisions are, good combat outcomes are the ultimate measure of the value of network-centric operations. An example application shows how these approaches can be combined. The mathematical approach is used to filter out preferred network and clustering assumptions, which are then tested in a simulation environment. This allows the development of both network-based Measures of Command and Control Effectiveness and higher-level Measures of Force Effectiveness.

Acknowledgments

The authors wish to express their gratitude to several individuals who provided guidance and assistance to the project. In the United Kingdom, we thank Lynda Sharp and Tim Gardener (Dstl), Lorraine Dodd (Qinetiq), and Christopher Watson (BAE Systems) for their technical support and continued interest in the application of the methodologies presented in this report. We also thank the Controller of Her Britannic Majesty's Stationery Office for granting permission to publish the Crown Copyright material contained in the text, especially the discussion of the Rapid Planning Process found in Appendix A. At RAND, we thank Gina Kingston, a visiting fellow from the Defence Science and Technology Organisation, Australia, for her helpful comments and early review of the text, and Tom Sullivan for his suggestions concerning the use of the mean squared error to inform knowledge. We also thank R. J. Briggs for his assistance in developing some of the code that was used to apply the concepts. Christopher Pernin (previously on secondment to Dstl) provided the analysis in Appendix C and contributed extensively to developing the Collaboration Metric Model used to illustrate the value of the metrics. We also thank former RAND colleague Jimmie McEver and Roger Forder at Dstl for their careful reviews of this work. Their comments and suggestions strengthened the finished report. Finally, we thank Robin Davis for her assistance in preparing this document for publication.

Abbreviations and Glossary of Terms

AA Bde	Air Assault Brigade
Accuracy	The degree to which information agrees with ground truth
ACP	Ammunition Control Point
AH Regt	Attack Helicopter Regiment
Armd Bde	Armoured Brigade
Armd Div	Armoured Division
Awareness	A realisation of the current situation
Bias	Error in an estimate introduced by systematic distortions
BSA	Brigade Supply Area
C4ISR	command, control, communications, computers, intelligence, surveillance, and reconnaissance
CEC	Cooperative Engagement Capability; a capability that combines data from all platforms in an operation and allows the combined data to produce a better shared CROP
CEP	circular error probable
Cluster	A set of network nodes possessing full shared awareness
CMM	Collaboration Metric Model
CoA	course of action

Collaboration	A process in which operational entities actively share information while working together towards a common goal
Complexity	The condition of having several interrelated parts in a network with several interrelated operational entities. Kolmogorov definition: The length of the shortest binary program needed to compute a string of data; the minimal description length
Conceptual space	The conceptual space of a commander is the space defined by the values of his critical information requirements
CROP	common relevant operating picture; a view of the battlespace shared by all friendly forces
DLM	dynamic linear model
DSA	Divisional Supply Area
Dstl	Defence Science and Technology Laboratory
FOB	Forward Operating Base
FSG	Forward Support Group
Full shared awareness	A set of network nodes that (1) share information, (2) agree on the same set of critical information elements, and (3) agree on the current values of the agreed critical information elements
Information entropy	A measure of the average amount of information in a probability distribution (also referred to as Shannon entropy)
Information superiority	The ability to collect, process, and disseminate information as needed; anticipate changes in the enemy's information needs; and deny the enemy the ability to do the same
IPB	intelligence preparation of the battlefield
Knowledge	Accumulated and processed information wherein conclusions are drawn from patterns
Logically connected nodes	Nodes with a communication *path* between them

MADM	multiple attribute decisionmaking
Measures	Standards for comparison
Mech Bde	Mechanised Brigade
Metrics	Mathematical expressions that evaluate both the relative effect of alternatives and the degree to which one is better or worse than another
MLRS Regt	Multiple-Launch Rocket System Regiment
MSE	mean square error; a measure of the accuracy of an estimate. It is the sum of the bias and the precision of the estimate
Mutual information	The amount of information gained about random variable X based on information gained about dependent variable Y
NAI	named area of interest
PCPR	perceived combat power ratio
Physically connected nodes	Nodes with a communications *link* between them
Plecticity	The ability of a connected set of actors to operate synergistically via the connectivity among them
Precedence weighting	A multi-attribute decisionmaking method
Precision	The degree to which multiple observations are close together
RPD	Recognition Primed Decision
SA	situation awareness
SAW	simple additive weights; a multi-attribute decisionmaking method
Shared awareness	The ability of a decisionmaking team to share realisations
TAI	target area of interest

CHAPTER ONE

Introduction

New information technologies introduced into military operations provide the impetus to explore alternative operating procedures and command structures. New concepts such as network-centric operations and distributed and decentralised command and control have been suggested as technologically enabled replacements for platform-centric operations and centralised command and control. As attractive as these innovations may seem, it is important that military planners responsibly test these concepts before their adoption. To do this, models, simulations, exercises, and experiments are necessary.

Objective

The major objective of this work is to produce a method to assess the effects of information gathering and sharing across an information network on the quality of decisions taken by a group of local decisionmaking elements (parts of, or a complete, headquarters). The effect is measured in terms of the reduction in uncertainty about the information elements deemed critical to the decisions to be taken at these local decisionmaking elements. We are thus assuming that the set of information elements necessary to produce a local conceptual picture of the battlespace is known.[1] The issue here is the degree of

[1] Other experimentally based research work in the United Kingdom is considering what these factors are in different scenarios.

confidence with which they are known, as measured by the local decisionmaking element's level of knowledge.

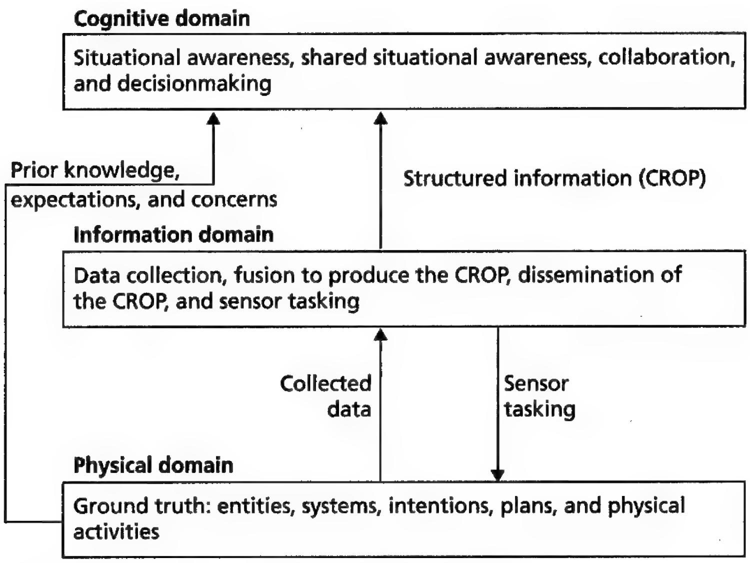
The term 'knowledge' has several meanings, and therefore it is important that, at the outset, we define what it means in the context of the decisionmaking processes described in this work. Formally, we define *knowledge* to be accumulated and processed information wherein conclusions are drawn from patterns. Information elements accumulated over time form patterns that can be matched to known patterns. The more reports confirming a given pattern, the less uncertainty remains and the more knowledge is gained.

The Information Superiority Reference Model

In terms of the categorisation developed by Alberts et al. (2001), we are representing the flow of information about the *physical domain* around the network in the *information domain* and its effect (in terms of knowledge, situation assessment, shared awareness, and decisionmaking) in the *cognitive domain.* These concepts are embodied in the *information superiority reference model* depicted in Figure 1.1. Information superiority is a term used to express the ability of one side in a conflict to impose its will over the other based on superior information collection, processing, and dissemination capabilities. Formally, we define *information superiority* to be the ability to collect, process, and disseminate information as needed; anticipate changes in the enemy's information needs; and deny the enemy the ability to do the same.

Both sides in a conflict generally have different perceptions of a single reality, referred to as the *situation.* Figure 1.1 shows how the three domains contribute to this perception. We list the major activities performed in each of the domains in each of the boxes. The physical domain is where reality, or ground truth, resides. In addition to physical objects, such as weapon systems, terrain features, and sensors, this domain also contains intangibles, such as enemy intent, plans, and current and projected activities. A complete assessment of the situation will contain estimates about each.

Figure 1.1
The Information Superiority Reference Model



RAND *MG226-1.1*

In the information domain, data are extracted from the physical domain and processed to form structured information in the form of a common relevant operating picture (CROP). Three primary functions are performed in the information domain: collecting data through the use of sensors and sources, including tasking sensors, to close gaps in the data; processing the data through the fusion process to produce the CROP; and disseminating relevant parts of the CROP to friendly units. The last step contributes to the collaboration process in the cognitive domain in which the shared CROP is transformed into a shared awareness of the current and future situation that can be used to gain understanding of threats and opportunities as well as the subsequent decisionmaking regarding an appropriate course of action. Our quantified assessment of the difference due to

local collaboration is a knowledge-based metric and hence resides in the cognitive domain.[2]

Finally, the human activities associated with using the information available to form an estimate of the situation are accomplished in the cognitive domain. To the extent that decisionmaking teams exist, they collaborate to form a level of situational awareness. In addition to the CROP produced in the information domain, individual team members and the decisionmaker may have prior information from processes like the intelligence preparation of the battlefield (IPB) available to support their deliberations. Finally, the decisionmaker is likely to have concerns and expectations about the performance of his own forces, as well as the enemy's, that would colour his assessment of the situation and therefore his decision. These elements are depicted in Figure 1.1 as emanating directly from the physical domain.

This report documents the mathematical constructs and metrics used to assess the effectiveness of the various operating schemes and command arrangements.

Research Approach

The basis of our approach is to bring together two sets of ideas, which have been developed thus far from rather different perspectives. The first of these comes from the work performed as part of a project on command and control in operational analysis models within the UK Ministry of Defence Corporate Research Programme. The programme aims to provide the Ministry of Defence with the ability to carry out fundamental research not tied to particular procurement programmes. In this case, it has led to the development of the Rapid Planning Process (Moffat, 2002) as a construct for representation of the decisionmaking of military commanders working within stressful and fast-changing circumstances. The process is now well accepted and has been included in a number of key command and

---

[2] *Collaboration* in this context is taken to be a process in which operational entities actively share information while working together towards a common goal.

control–centred simulation models developed or under development by the UK Defence Science and Technology Laboratory (Dstl). Such a representation approximates to the 'simple decisionmaking' of Alberts et al. (2001) in which the information elements and the criteria for decision are known and a satisficing strategy is adopted.

The second set of ideas comes from the work on modelling the effects of network-centric warfare, carried out recently by the RAND Corporation for the US Navy (Perry et al., 2002). In this work, the effects of collaboration across alternative information network structures in prosecuting a time-critical task can be assessed using a spreadsheet model. The benefits and costs of local collaboration are quantified using a relationship based on information entropy as a measure of local network knowledge. The effects of network complexity and the completeness of the information collected are also reflected in the overall assessment of the quality of the information made available to the decisionmakers.

To merge these two ideas, we examine the decisionmaking process among networked headquarters. We postulate that improved decisions are contingent on increased knowledge and, therefore, on diminished uncertainty. The pattern-matching features of the Klein Recognition Primed Decision (RPD) model (Klein, 1989) are used to match current estimates of critical information elements to the decisionmaker's set of typical situations or internalised patterns. A match is made when the current estimates lie within the comfort zone of one of the typical situations.

There are several analytic techniques available that are able to match estimates of values of multiple information elements to sets of typical situations—variously referred to as pattern-matching techniques or classification processes. In this work, we rely on the matching algorithms within the Rapid Planning Process mentioned earlier and discussed in detail in Appendix A. The decision to be taken in this case is the selection of an appropriate course of action based on the closeness of the current critical information element estimates to one of the typical situations.

Since, in a military operation, a rapid decision is usually desirable, the focus now centres on the means used to collect information

about the uncertain critical information elements, the ease with which this information is shared among participants in the operation, the quality of the resulting processed information, and its effect on knowledge. The methodology then turns to examining the structure of the decision networks and the quality and quantity of the collaboration exercised on it, and how both contribute to overall knowledge and, by extension, better decisions.

Organisation of This Report

In the next chapter, we set forth the framework for thinking about decisionmaking in a network. In Chapter Three, we address the uncertainties associated with information elements needed to support decisions, and suggest statistical representations that include a knowledge metric. Chapter Four examines the effects of collaboration on network performance. In Chapter Five, we explore the effects of network complexity on network performance and combine collaboration and network complexity to achieve a single metric that measures the performance of networked clusters of decision nodes.

In addition, we include three appendixes. Appendix A describes the Rapid Planning Process, and Appendix B discusses information entropy used in the development of the knowledge metric. Finally, Appendix C describes an application of the measures and metrics discussed in the text to the logistics command and control problem discussed in Chapter Two. Appendix C also discusses how the Measures of Command and Control Effectiveness, examined in the main body of this report, may be combined with combat models to assess the effects of increased knowledge on force effectiveness.

CHAPTER TWO

Decisions in a Network

Western militaries are formulating new visions, strategies, and concepts that rely on acute situational awareness, the transformation of information into knowledge, and rapid, secure means of sharing knowledge. They seem to be placing great reliance on networked forces that are fully integrated with joint, national, and coalition or allied systems. To achieve these goals, militaries must create and leverage information superiority. It is foreseen that, under some circumstances, a mix of command and control capabilities would be integrated with weapon systems and forces on an end-to-end basis through a network-centric information environment to achieve significant improvements in awareness, shared awareness, and collaboration (Alberts et al., 2001; Alberts et al., 2002).

The ultimate effect, however, is on the quality of the decision-making process and the decision itself. These decisions ultimately lead to actions that change the battlespace. In this report, we are thus concerned with the quality of these decisions, i.e., the planned outcome, rather than the effect in the physical domain. It is almost an article of faith that a richly connected network of decision nodes will perform better by improving the quality of decisions. However, we need to quantify this benefit as well as consider and quantify the downside of such information sharing (such as the effect of information overload and the problems associated with resolving disconfirming evidence).

The Decision Model

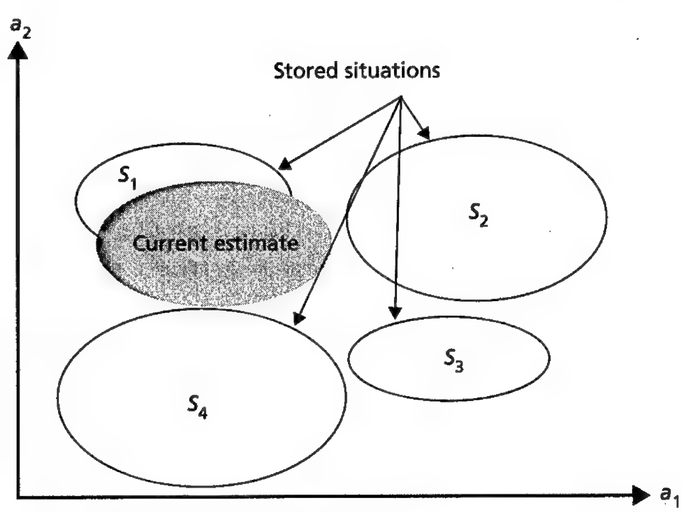
For this work, we assume that the decisions taken by the various decision elements in the headquarters network are selections of courses of action (CoAs) in response to the perceived situations. The CoAs prescribe actions to be taken in the event that the situation in the battlespace deviates from what is expected. Ideally, a mutually exclusive and collectively exhaustive set of CoAs is known to the decisionmakers, and all they need do is collect information that informs the perceived situation. In general, this is only partially true: CoAs can also be developed in response to unfolding events—events that may not have been perceived a priori. However, it is a reasonable assumption when representing expert decisionmakers in stressful and time-critical circumstances.

This approach is consistent with the naturalistic decisionmaking paradigm of the RPD model, introduced by Gary Klein (1989). Klein argues that experienced decisionmakers store up a set of typical situations and responses over time. They search the environment for clues, cues, and expectancies that might clarify the situation. Once the situation is perceived to match one of their stored situations, the decisionmakers are then able to respond accordingly by selecting what they feel is an appropriate course of action—generally something that has worked in the past. However, if the situation is not clarified, they seek additional information or examine the situation to determine causes for the lack of clarity. This assessment could lead to the modification of a typical situation and response or to the creation of a whole new stored experience. The latter behaviour results in the creation of new CoAs.

Matching the current situation to one of the decisionmaker's stored situations is clearly a subjective process. Each decisionmaker assesses the current values of what are considered to be critical information elements and decides whether the values are 'close enough' to one of the stored situations. The choice of a 'good enough' stored situation defines what we refer to as the decisionmaker's *comfort zone.*

Figure 2.1 illustrates what is going on.[1] In this case, the commander's conceptual space is described in terms of two critical information elements, a_1 and a_2. The ground-truth values of these information elements are not known with certainty and therefore are considered to be random variables with known densities. The ellipses in the diagram represent the decisionmaker's comfort zones for each of the stored situations. The centre of each is the desired value set, and the major and minor axes represent acceptable deviations from this desired set. Both the centre and the axis lengths in each direction are fixed. The centre of the shaded ellipse represents the current estimates for a_1 and a_2, and the axes represent the uncertainty in the estimate based on the covariance between the two.

Figure 2.1
Decisionmaker's Conceptual Space and Stored Situations



NOTE: Adapted from Moffat (2002), p. 45.
RAND *MG226-2.1*

[1] See Moffat (2002) for a more complete discussion.

In the diagram, we depict four stored situations, each with its degree of acceptable uncertainty depicted by the size of its ellipse. The shaded ellipse is the current estimate, and its size represents the uncertainty in the estimate. In this case, although the estimate is closest to situation S_1, it does not fall completely in the comfort zone. The issue then is to discern how close the shape must be to declare a match. In practise, this is a subjective process dependent in part on the decisionmaker's attitude to risk.

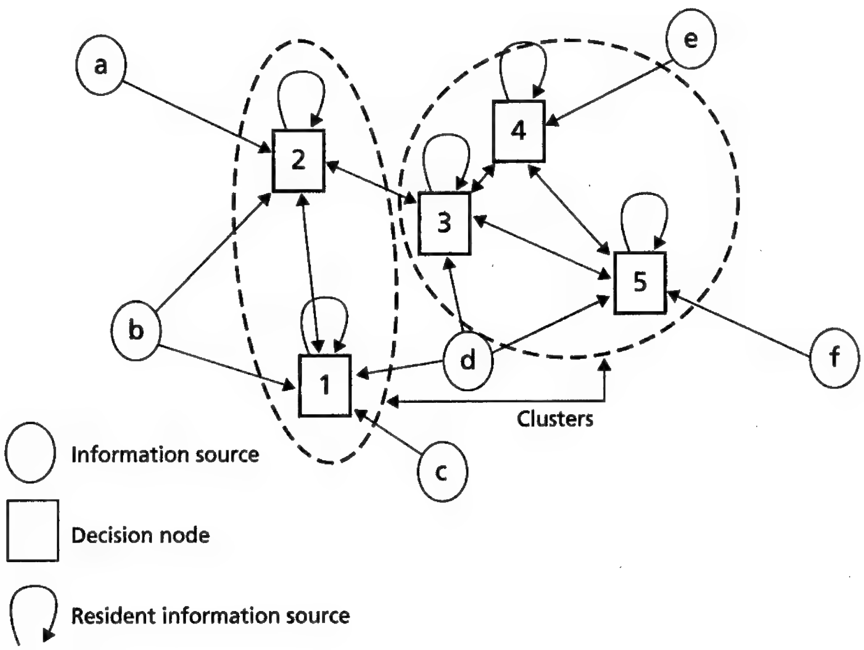
Estimators

Through observations of the battlespace, sensors and other information sources generate estimates for the information elements deemed critical to the decision. As we discuss in the next chapter, the uncertainty associated with the information elements is expressed in terms of probability distributions, the means of which are estimates of the ground-truth values. The quality of the estimates is therefore of concern to us as we assess the contribution of networking to the quality of the decisions taken. The mean of a probability distribution being a parameter of the distribution, we turn naturally to parameter estimation theory to assess the quality of the information available to the decisionmaker, and we examine how the quality of the estimates contribute to knowledge. Mathematical constructs from estimation theory exist for the quality of estimates such as accuracy, bias, precision, sufficiency, efficiency, and consistency. We discuss some of these terms more fully in Chapter Four.

A Networked Decision Model

Figure 2.2 depicts a simple network of decisionmaking nodes that are connected to each other to form a decision network. In Alberts (2001), the point is made that such a node-based network should represent actors, decisionmakers (or knowledgeable entities), and sen-

Figure 2.2
Network of Decisionmaking Elements



RAND *MG226-2.2*

sors (in the most general sense of information gatherers). We put the focus here on information gathering and decisionmaking. Each node thus represents either a 'local decisionmaker'—i.e., a local commander who needs to make decisions, or an information source—i.e., a collection facility such as a sensor, a processing facility such as a fusion centre, or a source of information about future plans. Decisions are made based on the information available to them either locally, from collection assets and information processing facilities elsewhere in the network, or from other local decisionmakers with whom they are connected. The connectivity depicted is logical and not necessarily physical. The structure of this network (the local commanders represented and how they link up across a network) will

be determined by the way we choose to organise the system and develop a plan.

Information is thus available from three sources: other decision nodes, external information sources, and information resident at the decision node.[2] In this depiction, we are concerned solely with such information flows.

Clusters

In Figure 2.2, some of the *decision nodes* are linked together to form a *cluster* that allows for *local* sharing of information. The term 'cluster', as used here, refers to a set of network decision nodes that (1) share information, (2) agree on a common set of critical information elements, and (3) agree on the current value of the agreed critical information elements and degree of uncertainty associated with the current values. The term 'local' refers to proximity in terms of logical connectivity. It does not necessarily imply physical nearness. In addition, we assume that each of these clusters supports distributed decisionmaking over time. Hence, we consider the process to be dynamical.

Clusters of decision nodes have the following properties:

- Only decision nodes can be members of a cluster.
- A cluster forms a *complete* graph. All decision nodes communicate with each other, thus producing $n(n-1)$ connections, but these connections are not necessarily physical.
- All decision nodes in a cluster are *self-aware*. Each decision node is aware of its own status and is able to inform others in the cluster.
- A cluster could consist of a single decision node, a number of nodes, or perhaps even all decision nodes in the network.
- Clusters may or may not communicate with each other.

---

[2] Resident information is sometimes referred to as 'organic information'. This expression is the preferred term in the Office of Force Transformation's network-centric operations framework (Office of Force Transformation, Network-Centric Operations Conceptual Framework, Version 1.0, 12 April 2004; available at www.oft.osd.mil).

- A *decision network* consists of the union of clusters. The total network consists of the decision network and all supporting nondecision nodes.

For example, in Figure 2.2 decision nodes 1 and 2 share information and therefore form a cluster. Decision nodes 3, 4, and 5 also share information and therefore form another cluster. Note that although decision nodes 2 and 3 may share information with each other, neither shares information with the other decision nodes in the other's cluster. In the academic literature, 'small world networks' have taken an approach similar to this in which highly clustered sets of nodes are linked by longer-range 'shortcuts'. These types of links lead to desirable network properties such as a high clustering coefficient (a measure of how well the network is linked locally) and a low average path length (a measure of the mean number of links between two randomly chosen nodes).[3]

Partitioning

We mentioned earlier that our goal is to assess the degree to which networked headquarters increase (or decrease) the knowledge available to the decisionmakers and at what cost. We stop short of actually taking the decision but rather measure success on the premise that more knowledge improves decisionmaking.

One way to affect network knowledge may be to rearrange or *partition* the network clusters. In Figure 2.2, for example, there are several possible partitions, ranging from five separate independent clusters of a single decision node each to one cluster consisting of all five decision nodes.[4] The question therefore is how best to partition the network to improve knowledge at an acceptable cost.

---

[3] See Watts (1999) and Albert and Barabási (2002).

[4] For a three-decision node network, the number of partitions is five; for a four-decision node network, it increases to 15; and for the five-decision node network, depicted in Figure 2.2, the number of possible partitions is 49. The number of partitions for n nodes is

$$P_n = \sum_{k=1}^{n} S(n,k),$$

Requirements for a Model of the Process

We now take up the requirements for the general model in more detail. There are many ways in which networks can be evaluated, using a variety of methods such as petri nets, Bayesian networks, or Neural Nets. The approach chosen depends on the particular characteristics of the network and the metrics that have analytic value. Following the ideas of Claude Shannon, we use *information entropy* as a key construct in developing metrics—since we wish to focus on information—and how it is converted into knowledge. We also use estimation theory to assess the quality of the estimates of the critical information elements needed to take decisions. In addition, we wish to capture the network dynamics of local information sharing, clustering in the form defined above, local collaboration, and the costs associated with complex network structures, since these capture core aspects of potential future headquarters structures. It is for these reasons that we have adopted the method presented here.

Consider one of these clusters, i. Suppose a local decisionmaker within cluster i must take a critical decision at time t. Estimates of the information required for the cluster to render a decision is accumulated over time so that if

$$\mathbf{x}_i(t) = \left[x_{i,1}(t), x_{i,2}(t), \cdots, x_{i,C}(t)\right]$$

represents the current estimated values for the C cluster-agreed critical information elements $\{a_1,\ldots\ldots,a_C\}$ needed at time t, the historical matrix of values for the estimates of the critical information elements is represented by the $t \times C$ matrix:

$$\left[\mathbf{x}_i(1), \mathbf{x}_i(2), \cdots, \mathbf{x}_i(t)\right] = \left[x_{i,k}(j)\right]_{t\times C} = \begin{bmatrix} x_{i,1}(1) & x_{i,2}(1) & \cdots & x_{i,C}(1) \\ x_{i,1}(2) & x_{i,2}(2) & \cdots & x_{i,C}(2) \\ \vdots & \vdots & \ddots & \vdots \\ x_{i,1}(t) & x_{i,2}(t) & \cdots & x_{i,C}(t) \end{bmatrix}.$$

---

where $S(n,k)$ (also known as the Stirling number) is the number of partitions of n nodes into k nonempty sets and $S(n,k) = S(n-1,k-1) + kS(n-1,k)$ (Jackson and Thoro, 1989).

Each element in the matrix represents the perceived value (estimates) of the critical information element, a_k, at time j for cluster i.

We wish to represent the local decisionmaking process within cluster i, using ideas based on the Rapid Planning Process. We thus represent the local conceptual space of the decisionmakers within cluster i by a space spanned by a small number, C, of information elements that are the key concerns of the decisionmakers within the cluster, as depicted for $C = 2$ in Figure 2.1.

Framing

For the entire network, we assume there is a maximum of N of these critical information elements, a_k, and therefore $\mathbf{A} = \{a_1, \ldots\ldots, a_N\}$ is the global set of critical information elements that we shall refer to as the *superset of critical information elements*. Each of the critical information elements, a_k, is perceived to have the value $x_{i,k}(j)$ at time step j within cluster i.

Suppose $N = 4$, so that the complete information set is $\mathbf{A} = \{a_1, a_2, a_3, a_4\}$. For each cluster, the local conceptual picture will be 'framed' by selecting a subset of **A**. For example, the local conceptual space of a cluster might be framed by the set of elements $\mathbf{A}_1 = \{a_1, a_2\}$. The space of a second cluster might be framed by the set $\mathbf{A}_2 = \{a_2, a_3, a_4\}$. Then, given that the two clusters collaborate, the local collaboration between them results in a shared conceptual space that is framed by the elements

$$\mathbf{A}_1 \cup \mathbf{A}_2 = \{a_1, a_2\} \cup \{a_2, a_3, a_4\} = \{a_1, a_2, a_3, a_4\} = \mathbf{A}_{1,2}.$$[5]

Shared Awareness and Clustering

A cluster of decision nodes as defined earlier corresponds to a form of *shared awareness* if the information shared among the cluster nodes is available and internalised at each decision node in the cluster. By 'shared awareness', we mean the ability of the decision nodes in the

[5] In this simple example, we have that $\mathbf{A}_{1,2} = \mathbf{A}$; however, this is not always true.

cluster to share realisations about the critical information elements. We further state that the nodes of a cluster possess *full shared awareness* if, in addition to sharing the same set of critical information elements, they further agree on the values each of these should take at a given time.[6]

These perceived values, $x_{i,k}(j)$, of the critical information elements form the input data to cluster i at time step j as described in Appendix A (The Rapid Planning Process) at stage 1 (observation analysis and parameter estimation). Within cluster i, we assume there are a shared number of fixed patterns or stored situations in the shared local *conceptual space* that are the areas of the space about which decisionmakers within the cluster are particularly concerned. These are represented by multivariate normal probability distributions in the conceptual space in the basic approach, as described in Appendix A. However, when a multivariate normal representation is not appropriate, more general methods must be applied, as will be discussed later. In either case, these fixed patterns are assumed to be directly linked to one of a small set of key courses of action (or missions) from which the local decisionmakers within the cluster can choose.

A Simple Logistics Example

Sustainment of deployed forces is one of the more difficult logistics tasks. In this simple model, we do not claim to have examined all the problems associated with supplying the force. In fact, we explore a single decision: allocating supplies to competing friendly units. This would be part of a sustainment plan, and our task is to examine how various decisionmakers contribute to the plan across a simple network of information sharing.

Figure 2.3 depicts the structure of a push (a) and pull (b) system for logistics resupply from a holding point to two local commanders.

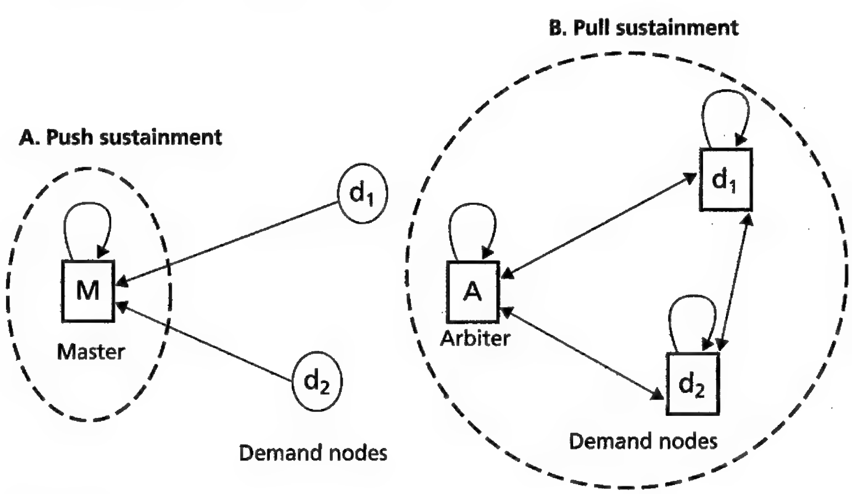
[6] This relates to the models of situational awareness such as those discussed in Endsley (1995) and Feltham, Sheppard, and Cooper Chapman (2003).

The allocation decision is made by the master in Figure 2.3a and the arbiter in Figure 2.3b.

In Figure 2.3a, the master node decides which local commander has priority, and therefore there is no benefit to be gained from the two demand nodes collaborating. As a result, the demand nodes (local commanders) are considered sources of information about their own stock levels so that the critical information set required by the master is $A=\{a_1,a_2\}$, where the numerical subscripts refer to the supply levels at the two demand points. Consequently, the network consists of the decision node 'cluster' and the two demand nodes. Information about global stock levels is only available at the master node.

In Figure 2.3b, the arbiter responds to the demands from the local commanders. All three nodes in this case are decision nodes and require the same information to make their decisions—the local supply levels $A=\{a_1,a_2\}$, as in the master case. The local commanders place demands on the arbiter based on their anticipated requirements, and the arbiter allocates stocks based on knowledge of global supplies

Figure 2.3
Networked Sustainment Decisions



RAND *MG226-2.3*

and anticipated future needs at both demand nodes. The knowledge of global stock levels is based on shared information from the demanding nodes and the stock levels available to the arbiter. In this case, we can consider the benefit of the local commanders collaborating in order to ensure that their demands are placed by taking account of global knowledge about stock levels. The network therefore consists of a single cluster that comprises the three decision nodes, as depicted in the diagram.[7]

In the push case, no other partitions are possible because there is only one decision node. In the pull case, however, it is possible to consider the arbiter and one of the headquarters nodes in a single collaborating cluster and the other a single decision node cluster. Operationally, we would expect the arbiter, in this case, to give priority to the connected headquarters, with the residual supply going to the single-node cluster. It is not possible, however, to partition the two headquarters as a single cluster with the arbiter as a single decision node cluster because it would violate the operational concept. For this example, we consider the single cluster in each case.

Each cluster supports local decisionmaking within the cluster. We can enrich the representation by adding an information node that supplies more detail on the operational plan, the synchronisation matrix of the forces, and the resultant likely pattern of demand for stocks. We focus here on the demand for fuel supplies to make the example more concrete. In Chapter Four, we will discuss the implications of these two modes of supply in terms of information sharing through collaboration and network knowledge. Later, we will address the costs of achieving this level of knowledge as well.

[7] All three agree that the local and global stock levels are the critical information elements, they all share information about the value of these critical information elements, and they all agree on these values.

CHAPTER THREE

Representing Uncertainty

The decisions within each of the clusters must be taken, in most cases, without full knowledge of the values of the critical information elements needed to support the decisions. The degree of uncertainty depends on the information collection assets devoted to the cluster's critical elements of information and the extent to which collaboration among the cluster decision nodes is facilitated. *Information entropy* is a reasonable estimate of the uncertainty, and consequently we use a probabilistic entropy model to represent the uncertainty associated with the critical information elements needed within the cluster. The reports on the values of the critical information elements are treated as estimates of the means of the distributions describing their uncertainty, and therefore the quality of the estimates is assessed using concepts from estimation theory.

Decisions

The decisions taken within each of the clusters depend on the scenario represented. In Figure 2.3, we depicted a simple logistics example in which the decision is the quantity of supply to allocate to each demanding unit. In general, we focus on operational and tactical decisions made at the division/brigade, ship group, or equivalent level and below. The decision taken within the cluster depends on the current estimated values of the critical information elements.

The dependency among the critical information elements is modelled using the Rapid Planning Process, provided that uncertainty can be expressed in the form of normal distributions. The Rapid Planning Process is a set of algorithms that together represent local command decisionmaking at each of the decision nodes (Moffat, 2002). The process requires that the commander's local 'conceptual space' be spanned by a small number of critical information elements. These elements are a subset of the total set, $\{a_1,\ldots\ldots,a_N\}$, of information elements considered across the network. In the basic formulation, a dynamic linear model (DLM; see Appendix A) is then used to represent the decisionmaker's estimates of the values of these factors over time. Ideally, through this process, additional information from collection assets or from collaborating elements in the network serves to reduce uncertainty and therefore increase understanding.

A Multivariate Normal Model

We begin first with a simple case in which we assume that the uncertainty in these critical information elements is represented by a multivariate normal distribution, and we are considering all the information elements $\mathbf{A}=\{a_1,\cdots,a_C\}$ shared across a cluster.[1] Their values are represented by the random vector $\mathbf{X}=[x_1,x_2,\cdots,x_C]^T$. In this case, the DLM can be used to make a local assessment of the overall uncertainty of the vector of critical information elements within the cluster. The uncertainty in the vector is represented as the multivariate normal distribution

$$f(\mathrm{X})=\frac{1}{\sqrt{(2\pi)^C|\Sigma|}}e^{\left(-\frac{1}{2}[\mathrm{X}-\boldsymbol{\mu}]^T\Sigma^{-1}[\mathrm{X}-\boldsymbol{\mu}]\right)},$$

[1] We will deal later with the more general case in which this assumption need not hold.

where $\mu = [\mu_1, \mu_2, \cdots, \mu_C]$ is the mean and

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \Sigma_{1,2} & \cdots & \Sigma_{1,C} \\ \Sigma_{2,1} & \sigma_2^2 & \cdots & \Sigma_{2,C} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{C,1} & \Sigma_{C,2} & \cdots & \sigma_C^2 \end{bmatrix}$$

is the covariance matrix. The off-diagonal elements are the covariance values between the random variables x_i and x_j, calculated as $\Sigma_{i,j} = \mathrm{E}(x_i - \mu_i)(x_j - \mu_j)$. The value

$$\rho_{i,j} = \frac{\Sigma_{i,j}}{\sigma_i \sigma_j}$$

is the correlation between the random variables, x_i and x_j. When $i = j$, then $\Sigma_{i,j}$ is just the variance σ_i^2 depicted along the diagonal in the covariance matrix. The entropy of the distribution (as we will discuss in more detail later) can be easily calculated from the covariance matrix and is then used as the basis for a quantifiable metric of the knowledge available to the cluster. With improved knowledge, we can refine the estimates of the critical information elements to more closely reflect ground truth.

Knowledge from Entropy

Decisions taken within a cluster depend on the degree to which the local decisionmakers know the true values for each of the critical information elements. $f(\mathrm{X})$ represents the level of uncertainty associated with the values of the critical information elements. It therefore forms the basis for measuring the level of knowledge. To quantify the level of knowledge, we apply the concept of information entropy, borrowed from information theory.

Information entropy, sometimes referred to as Shannon entropy, measures the amount of information in a probability distribution

(Shannon, 1948). Shannon entropy for a probability density function, $f(\mathrm{X})$, is defined to be the expected value of the negative logarithm of $f(\mathrm{X})$, or

$$H(\mathrm{X}) = E[-\log f(\mathrm{X})] = -\int_{x_1}\int_{x_2}\cdots\int_{x_C} f(\mathrm{X})\log f(\mathrm{X})dx_C\cdots dx_2 dx_1.$$

If, as in this case, $f(\mathrm{X})$ is continuous, $H(\mathrm{X})$ is referred to as *differential entropy*.[2]

For the multivariate normal distribution, the differential entropy is calculated to be

$$H(\mathrm{X}) = \frac{1}{2}\log(2\pi)^C|\mathbf{\Sigma}| + \frac{C}{2} = \frac{1}{2}\log\left[(2\pi e)^C|\mathbf{\Sigma}|\right], \tag{3.1}$$

where $|\mathbf{\Sigma}|$ is the modulus of the determinant of the covariance matrix $\mathbf{\Sigma}$ and $C \le N$ is the number of information elements critical to the cluster.

In this work, we are interested in relative entropy, and therefore noting that $H(\mathrm{X})$ varies solely with the covariance (since C is held constant), we simplify equation (3.1) to $H_r(\mathrm{X}) = \log|\mathbf{\Sigma}|$. $H_r(\mathrm{X})$ is then a local measure of the (relative) information entropy. From now on, we will drop the subscript r.

Knowledge

Knowledge derived from entropy is a quantity, $0 \le K(\mathrm{X}) \le 1$, that reflects the degree to which the local decisionmakers within a cluster know the true values of the information elements, $\{a_1,\cdots,a_C\}$, and their interaction. For $K(\mathrm{X}) \to 1$, knowledge is considered to be good, and for $K(\mathrm{X}) \to 0$, it is considered to be poor.

For the multivariate normal distribution, $K(\mathrm{X})$ is calculated as follows. We first assume the existence of a maximum joint entropy,

$$H_{\max}(\mathrm{X}) = \log|\mathbf{\Sigma}|_{\max}.$$

[2] See Appendix B for a more detailed discussion of information entropy.

Physically, this can be interpreted to be the maximum uncertainty in the probability distribution, $f(\mathrm{X})$. If, for example, the information elements consist of the x- and y-coordinates associated with the location of an enemy unit, the maximum entropy might be associated with search area. Search areas are derivative of 'named areas of interest' (NAIs) or 'target areas of interest' (TAIs). If, through the IPB process, we are able to describe a circular area in an NAI or TAI within which we are virtually certain the enemy unit is located, we can then relate this information through a circular error probable (CEP) to the variance of the location in the x- and y-directions.[3]

If the maximum entropy is taken to be $\log|\boldsymbol{\Sigma}|_{\max}$, then the residual entropy at any given time is $\log|\boldsymbol{\Sigma}|_{\max} - H(\mathrm{X})$. A formulation for $K(\mathrm{X})$ therefore ensures that a value confined to the interval [0,1] is

$$
\begin{aligned}
K(\mathrm{X}) &= 1 - e^{-\left[\log|\boldsymbol{\Sigma}|_{\max} - H(\mathrm{X})\right]} \\
&= 1 - \frac{|\boldsymbol{\Sigma}|}{|\boldsymbol{\Sigma}|_{\max}}
\end{aligned}
$$

When the modulus of the determinant of the covariance is close to its maximum, knowledge is at a minimum, whereas for small values of the covariance determinant, knowledge is greatest.[4]

The Effects of Knowledge

As a basis for our consideration of the effects of knowledge, we use the domain structure depicted in Figure 1.1. We are principally concerned with the information and cognitive domains. Information derived from sensors or other information gathering resides in the information domain. It is then transformed into awareness and knowledge in the cognitive domain and forms the basis of decision-

[3] For unit location, if we assume that $\sigma_x = \sigma_y = \sigma$ and that $\sigma_{xy} = \sigma_{yx} = 0$, then the CEP is related to the common variance as follows: $\sigma = CEP/1.1774$. CEP in this formula is the radius of the maximum search area. See Burington and May (1958).

[4] See Perry et al. (2002).

making. Our metrics quantify this process through the use of information entropy and the derivative knowledge measures. Information sharing among nodes ideally tends to lower information entropy (and hence increases knowledge) because of the reduction in variance and the buildup of correlations among the critical information elements.

One of the key aspects of increased knowledge (and, correspondingly, reduced entropy) is thus an increased understanding of the correlations between variables. This means information can be gained about one critical information element (e.g., missile type) from another (e.g., missile speed). Such cross coupling is a key aspect for consideration as we extend our analysis from normal to more arbitrary probability distributions.

More General Models

The multivariate normal assumption is likely to be restrictive for some applications. The example above in which it was used to represent the location of a target is perhaps the best-known military application. A more general model for a cluster recognises that the uncertainty associated with each of the critical information elements is likely to be represented by unique probability distributions and that their joint distribution is either unknown or can be discerned only through a laborious combinatorial process.

For example, suppose we wish to model a US carrier battle group executing a cruise missile defence mission with its attached Aegis cruisers.[5] Our cluster in this case might consist of the decisionmakers on board each of the Aegis cruisers taking part in the mission—assuming that all commanders in the cluster are able to

[5] This is a very real problem examined extensively by the Royal Navy and the US Navy. In the United States, the Cooperative Engagement Capability (CEC) is being developed in response to the challenges of littoral warfare, the shrinking size of US and Allied navies, and improvements to adversary capabilities. CEC is an approach to air defence that allows combat systems to rapidly share unfiltered sensor measurement data and track data to enable a carrier battle group to operate collectively. See 'The Cooperative Engagement Capability' (1995).

share information with each other. The decision to be taken is when and where to engage an incoming enemy cruise missile.[6] We further assume that each weapon system (standard missiles on board the Aegis cruisers) requires the same information—the location of the target (latitude and longitude), its altitude and speed, its direction, and its type—so as to have a critical information set that is uniform among all decision nodes in the cluster:[7]

$$A = \{\text{location}, \text{altitude}, \text{speed}, \text{direction}, \text{missile type}\} = \{\mathbf{a}_1, a_2, a_3, a_4, a_5\}.$$

These are the information elements shared among the cluster, leading to full shared awareness within the cluster. The location of the missile has two components—latitude and longitude—and therefore we have $\mathbf{a}_1 = [a_{1,x}, a_{1,y}]$. The uncertainty of these components is taken to be bivariate normal as developed earlier. As tempting as it is to include altitude in location and model uncertainty in three dimensions, we recognise that altitude is bounded from below and therefore its uncertainty is better described using a density such as the lognormal or the gamma.[8] This situation is also true of speed. Direction, however, is circular and therefore bounded between 0 and 2π. If normalised on $[0,1]$, the uncertainty here can be represented by a beta density. Missile type is nominal, and therefore its distribution is empirical.

Although more realistic, this representation is clearly more problematic. Added to the complexity is the fact that not all the information elements are independent, and therefore their joint distribution is not likely to be multiplicative. For example, the speed of a missile is, in some part, a function of its type—as is its altitude. Its

[6] We omit a discussion of shooting policy, centralised versus decentralised command and control, and the participation of ground defence units. These are all interesting aspects of the problem and their examination in a network-centric environment will lead to the assessment of several alternative network structures and command and control arrangements—what our models are ultimately designed to accomplish.

[7] In this case, *direction* refers to the bearing of the incoming missile and not its inclination.

[8] Actually, it is likely bounded from above as well, and therefore one might argue for a beta distribution. In either case, a normal distribution is not appropriate.

location and direction at any point in time, however (ignoring its trajectory history), need not be.

Two problems arise from this more general representation: (1) describing the joint probability distribution, $f(\mathbf{A})$, needed to account for the dependencies within the critical information elements, and (2) combining the knowledge functions for each of the marginal distributions to create an overall measure of local knowledge. We discuss two methods for dealing with these issues: *multi-attribute assessment* and *mutual information.*

Multi-Attribute Assessment

The simplest (but perhaps not the most accurate) way to deal with the problem of combining the knowledge functions associated with multiple distributions is to create a weighted sum that represents the current level of knowledge of the combined critical information elements. Weights generally imply some notion of relative importance. Although indeed desirable, weights are not enough in all cases. What is needed is some way to represent the inherent dependencies among the information elements. Regardless of how well we are able to achieve this goal, a weighted sum is inherently flawed because of the fact that knowledge need not be additive. Nevertheless, as a means of comparison, the methodology has value.

The objective of multi-attribute assessment is to derive a single knowledge value that describes the joint level of knowledge about the critical information elements within a cluster and, ultimately, throughout the network. In the multivariate normal case, described earlier, this value is just the knowledge function derived from the distribution's information entropy. By deriving this single value, we model the assessment of a decisionmaker within the cluster, of the current estimates of the critical information based on information he has received, and his level of knowledge derived from these estimates. This, in turn, can be used to select a course of action (take a decision).

The two methods discussed here derive from multiple attribute decisionmaking (MADM) theory; in particular, the MADM techniques in which the decisionmaker is supplied with the value of different sub-attributes that contribute to an overall value. Generally, MADM methods are used when a decision must be made between two or more alternatives based on multiple attributes that have incommensurable units—for example, speed and direction. The choice of one technique over another depends on the nature of the attributes being combined and their relation to one another. Here we discuss two methods: simple additive weights (SAW) and weighted product.[9] In both methods, we use the terms 'system' and 'system instantiation' to refer to the combat situation at the time estimates of the critical information elements are to be assessed.

Simple Additive Weights Method

The SAW method (Fishburn, 1967) is perhaps the simplest method of aggregation and is a relatively old technique. It is cited in Article I, Section 2, of the US Constitution as a method to determine the degree of a state's representation in the Union.[10] It is generally used when the attributes are independent of each other. For a case in which there are C attributes shared across the cluster, we get

$$V(\mathrm{A}) = \Sigma_{i=1}^{C} \omega_i V(a_i),$$

where $V(\mathrm{A})$ is the value of the system instantiation with critical information elements, a_i. The term ω_i, $(\Sigma_{i=1}^{r} \omega_i = 1)$ is the weight (importance) of information element a_i, and $V(a_i)$ is its value (knowledge function in this case) for the instantiation being considered. Unfortunately, the likelihood that *all* information elements shared across the cluster are independent is very small, so this technique is not widely applicable except where the weights can be made to account for the dependencies in some way.

---

[9] For a complete discussion of several more methods, see Perry et al. (2002).

[10] See Yoon and Hwang (1995, p. 32).

Weighted Product Method

The weighted product method (Bridgman, 1922) is similar to the SAW technique, except in this case the values of the different attributes are multiplied. The general form of this method is

$$V(\mathrm{A}) = \prod_{i=1}^{C} \left[V(a_i)\right]^{\omega_i},$$

where $V(\mathrm{A})$, a_i, and ω_i are as above.

Although $V(\mathrm{A})$ might be used directly as a measure of combined system value, it is often the case that its value in relation to a positive ideal is used instead, so we obtain

$$\bar{V}(\mathrm{A}) = \frac{V(\mathrm{A})}{V(\mathrm{A}^*)},$$

where A^* is the positive ideal that may or may not be achievable.[11]

A similar approach is the Keeney-Raiffa multi-attribute utility method (de Neufville, 1990). In this method, the aggregation evaluation takes the form

$$\Omega V(\mathrm{A}) + 1 = \prod_{i=1}^{C} \left[\Omega \omega_i V(a_i) + 1\right],$$

where Ω is a normalising factor used to ensure consistency between the definition of $V(\mathrm{A})$ and the $V(a_i)$'s. The value of Ω is given by

$$\Omega + 1 = \prod_{i=1}^{C} \left[\Omega \omega_i + 1\right].$$

This technique is advantageous because it allows for the consideration of possible interactions between the attributes—something clearly desirable if we wish to account for dependencies. For example,

[11] The positive ideal case, also sometimes referred to as the most favourable case, is the instantiation with the highest overall value. The positive ideal case is selected from the existing instantiations, a combination of the existing instantiations or using the maximum possible value for each attribute.

if $C=2$, and $\{a_1,a_2\}$ is the set of information elements shared across the cluster, we get

$$V(\mathrm{A})=\omega_1 V(a_1)+\omega_2 V(a_2)+\Omega\omega_1\omega_2 V(a_1)V(a_2),$$

with $\Omega=(1-\omega_1-\omega_2)/(\omega_1\omega_2)$.

Precedence Weighting

Precedence weighting provides a method to get at the dependencies among the information elements. The weights are computed based on these dependencies. The relative importance of the information elements is assessed singly or in combination. For example, suppose we have decided that the information elements, shared across our CEC cluster, required to accurately engage an attacking cruise missile in our example are location, speed, direction, and missile type. Recall that the decision to be made is when and where to launch a standard missile to intercept the incoming enemy cruise missile. For each of the information elements and combinations of information elements, we ask: *Can the decision be taken with just this (these) information element(s)?* For example, can the decision to intercept be taken knowing only the location of the enemy cruise missile? with location and speed? etc.? For a given set of information elements of size r, 2^r, such questions must be asked. In this example, this amounts to 16 questions.[12] For large information sets, this method could become rather cumbersome, hence the omission of 'altitude'.

Other questions arise: If a decision can be made knowing the value of only one of the three information elements, what added value does knowing the other two provide? Are the information elements not used in the decision therefore still 'critical'? First, we assume that if an information element is designated as 'critical', it is needed to fully inform the decision. We recognise, however, that

[12] This includes the empty set, i.e., no information elements are available, and the entire set, i.e., all information elements are available. We would expect the answer to the former to be 'no' and the latter to be 'yes'. This is sometimes referred to as the information element *power set*.

decisions are taken with less-than-complete information but that there is some subset below which a decision is impossible or extremely risky—regardless of the urgency of the situation. In this example, having estimates for all information elements is better than two or one. One way to acknowledge this is to assign weights to various combinations of the information elements. However, doing this leads us back to subjective linear weighting. Consequently, we rely solely on counts for this method.

The answers to the questions determine the weights assigned to each element. Table 3.1 summarises the answers to the 16 questions.

The next step is to count the number of combinations that result in a 'yes' response in the last column of the table for each of the information elements. For example, of the 16 combinations here (including the empty set), location occurs in eight with a 'yes' response. In each of these, it was determined that a decision to engage the cruise missile could have been made with just the information elements in the combination. For the remaining three, the count is five each.

Because location alone was considered sufficient for a launch decision, any of the other combinations that included location were also considered sufficient. The other three information elements appear in exactly five 'yes' combinations because no two combinations of them were considered sufficient to order a launch but all three together were considered sufficient. If it were the case that each of the four information elements alone were sufficient to order a launch, then each would earn a score of 8, which is equivalent to equally weighted, independent information elements.

If all information elements were necessary and no lesser combination sufficient to launch, we get the same result. In this case, each information element would receive a score of 1.

Calculating the relative weights from these results consists of using the sum of the scores to normalise the weights so that

$$\omega_i = c_i / \sum_{j=1}^{C} c_j ,$$

Table 3.1
Precedent Weight Assessment

Location (a_1)	Speed (a_3)	Direction (a_4)	Type (a_5)	Yes/No
X	—	—	—	Yes
—	X	—	—	No
—	—	X	—	No
—	—	—	X	No
X	X	—	—	Yes
X	—	X	—	Yes
X	—	—	X	Yes
—	X	X	—	No
—	X	—	X	No
—	—	X	X	No
X	X	X	—	Yes
X	—	X	X	Yes
—	X	X	X	Yes
X	X	—	X	Yes
X	X	X	X	Yes

where c_i is the score for information element a_i. In this example, we would get the following weights: $\omega_1 = 0.348$ and $\omega_2 = \omega_3 = \omega_4 = 0.217$.

This method is practical only for small sets of information elements, since the dimension of the problem increases exponentially with the number of information elements. However, for most operational decisions, the number of critical information elements is small, and indeed, we assume this to be the case in this analysis. Even for those cases in which the number is large, it is likely that certain combinations are not worth examining because it is obvious that the combination would not be sufficient.

Mutual Information

Next, we discuss a more direct method to derive the multi-attribute knowledge function for a set of information elements shared across a cluster. Mutual information is derived from information entropy (see Appendix B) and deals directly with the issue of independence (or rather lack thereof) among the information elements. Although a joint probability density function is still required, mutual information

allows us to account for the dependencies even when the joint distribution is empirical.

What we desire is a mathematical construct that will allow us to modify our knowledge about a random variable (information element, X) based on our knowledge of a second random variable (information element, Y) when X and Y are not independent random variables. Because one random variable informs another, we refer to this construct as *mutual information.*

Relative Entropy

Relative entropy measures the 'distance' between two probability mass functions, denoted $D[p(x)\|q(x)]$. It is essentially the error incurred by assuming the true distribution for X is $q(x)$, when it is really $p(x)$. Relative entropy as defined by Cover and Thomas (1991)[13] is

$$D\left[p(x)\|q(x)\right] = \sum_{x \in X} p(x)\log\frac{p(x)}{q(x)}.$$

In this definition, we have

$$0\log\frac{0}{q(x)} = 0 \text{ and } p(x)\log\frac{p(x)}{0} = \infty.$$

If $p(x) = q(x)$, then $D[p(x)\|q(x)] = 0$. However, relative entropy is not a true distance metric because it is not commutative. That is,

$$D\left[p(x)\|q(x)\right] = D\left[q(x)\|p(x)\right]$$

is not always true.[14] Kullback (1978, p. 6) refers to the quantity

[13] See also Kullback (1978).

[14] A true metric satisfies the following properties:

A *metric space* is a pair (X, d), where X is a set and d is a *metric* on X (or a distance function on X), such that for all $x, y, z \in X$ we have:

$$D\left[p(x)\|q(x)\right]+D\left[q(x)\|p(x)\right]$$

as a measure of *divergence* between $p(x)$ and $q(x)$ and, therefore, a measure of the difficulty of *discriminating* between them.

Mutual Information

We use the concept of relative entropy to arrive at a measure of mutual information. Suppose we have two dependent random variables, X and Y, with joint probability mass function $p(x,y)$ and marginal mass functions $p(x)$ and $p(y)$. We define the mutual information to be the relative entropy between the joint mass function and the product of the marginal mass functions, or

$$\begin{aligned} I(X:Y) &= D\left[p(x,y)\|p(x)p(y)\right] \\ &= \sum_{x\in X}\sum_{y\in Y} p(x,y)\log\frac{p(x,y)}{p(x)p(y)}. \end{aligned}$$

Hence, $I(X:Y)$ defined in this way is the amount of information about X gained from Y.

Cruise Missile Type and Speed

Recalling our example again of the CEC cluster, we note that the type of enemy cruise missile threatening a friendly fleet can be inferred somewhat from its speed of approach. However, the relationship between the two is not exact because the missile may be operating at a speed other than its nominal speed and several of the missiles may operate at similar speeds. Nevertheless, if a report of missile

---

d is real-valued, finite and nonnegative.

$d(x,y)=0$ if and only if $x=y$.

$d(x,y)=d(y,x)$.

$d(x,y)\le d(x,z)+d(z,y)$.

(Taken from Kreyszig, 1978.)

speed is received, it is possible to improve our knowledge about the type of missile threatening us.

For example, suppose we let the random variables S and M represent the speed and enemy missile type, respectively. We define the joint probability mass function, $p(s,m)$, in Table 3.2.[15] Three missile types are listed as column headers. Continuous speed has been divided into four Mach intervals and are listed in the left-hand column. The entries in the table are the joint probability mass for the events $s_i \cap m_j$ or $p(s_i,m_j)$. The marginal distributions $p(s_i)$ and $p(m_j)$ are the probability that a missile is travelling within the range s_i and that the missile launched is of type m_j, respectively.

From this we calculate the mutual information:

$$I(M:S) = \sum_{j=1}^{3} \sum_{i=1}^{4} p(s_i,m_j) \log \frac{p(s_i,m_j)}{p(s_i)p(m_j)} = 0.222.$$

Table 3.2
Joint Probability Mass Function for Speed and Missile Type

	C601 (m_1)	C801 (m_2)	SS-N-27 (m_3)	$p(s_i)$
0–M0.75 (s_1)	0.05	0.04	0.20	0.29
M0.75–1.0 (s_2)	0.14	0.15	0.02	0.31
M1.0–2.0 (s_3)	0.03	0.05	0.07	0.15
>M2.0 (s_4)	0.04	0.01	0.20	0.25
$p(m_j)$	0.26	0.25	0.49	1.00

NOTE: The speeds are given in Mach units.

[15] Although it is always possible to create such a table, it is generally very difficult to obtain credible table entries. In most cases, a normalisation process is needed to convert whole numbers (generally from 1 to 10) supplied by operators to the joint probabilities. In any case, the entries are more likely to be subjective estimates.

Therefore, the amount of information about cruise missile type that can be gained from the speed of the incoming enemy missile is 0.222 nats.[16] Because $I(M:S)=I(S:M)$, we may also interpret this to be the amount of information about the speed of the incoming enemy cruise missile that can be gained from its type.

One way to use mutual information is to develop pairwise joint probability mass functions for all the critical information elements and calculate their mutual information. A high mutual information score between two information elements means that the two are non-randomly associated with each other, whereas a lower score signifies that the two are independent—that is, that the joint distribution of the two holds no more information than the information elements considered separately. Butte and Kohane (1999) use this pairwise assessment of mutual information to associate genes. They hypothesise that an association with high mutual information means that one gene is nonrandomly associated with another. They then select a threshold mutual information level and use the associations above the threshold to generate gene clusters or relevance networks.[17]

The next, and more difficult, step is to develop the appropriate weights from these pairwise associations. We have not addressed this problem as yet; however, it appears that the relevance network approach suggested by Butte and Kohane might be applicable.

Assuming a joint probability mass function can be found for all the information elements shared across a cluster, we can proceed as follows.

Entropy and Mutual Information

The knowledge function used to assess understanding relies on the calculation of information entropy. Consequently, it would be help-

---

[16] When information entropy is calculated using base 2 logarithms, the resulting measure of information present in the distribution is a 'nit'. When we use natural logarithms, the measure is the 'nat', with reference to the natural logarithm. See Appendix B for a fuller discussion.

[17] Two genes are connected in the network, if their mutual information scores exceed the threshold for that cluster.

ful to exploit the relationship between mutual information and information entropy. Fortunately, the relationship is quite straightforward. First, however, we need to develop the concept of conditional entropy.

Conditional entropies $H(X|Y)$ and $H(Y|X)$ are sometimes referred to as 'side information', i.e., the uncertainty (entropy) in one random variable is conditioned on another random variable.[18] If the random variables X and Y have a joint probability density, $p(x,y)$, the *conditional entropy* $H(X|Y)$ is defined as

$$\begin{aligned} H(X|Y) &= -\sum_{x\in X} p(x) \sum_{y\in Y} p(y|x) \log p(y|x) \\ &= -\sum_{x\in X} \sum_{y\in Y} p(x,y) \log p(y|x) \end{aligned}$$

From this, we can derive an expression for mutual information in terms of information entropy as follows:

$$\begin{aligned} I(X:Y) &= \sum_{x\in X} \sum_{y\in Y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} \\ &= \sum_{x\in X} \sum_{y\in Y} p(x,y) \log \frac{p(x|y)}{p(x)} \\ &= -\sum_{x\in X} \sum_{y\in Y} p(x,y) \log p(x) + \sum_{x\in X} \sum_{y\in Y} p(x,y) \log p(x|y) \\ &= -\sum_{x\in X} p(x) \log p(x) - \left[-\sum_{x\in X} \sum_{y\in Y} p(x,y) \log p(x|y) \right] \\ &= H(X) - H(X|Y). \end{aligned}$$

[18] In communications theory, the conditional entropies can be thought of in terms of a communications channel with input X and output Y. $H(X|Y)$ is then called the *equivocation* and corresponds to the uncertainty in the transmission from the receiver's point of view, and $H(Y|X)$ is called the *prevarication* and represents the uncertainty from the transmitter's point of view. Taken from Blahut (1987).

This is helpful because all quantities can be expressed in terms of information entropy.

The next step is to consider the multidimensional case. That is, how is the uncertainty in the perceived values of the critical cluster information elements $\{a_1,\cdots,a_C\}$ affected by knowledge of the value of information element y? By a simple extension of the relationship developed for two information elements, we get

$$I(X_1,X_2,\cdots X_C:Y)=H(X_1,X_2,\cdots X_C)-H(X_1,X_2,\cdots X_C\,|\,Y).$$

Assuming, of course, that the joint and conditional probabilities can be defined, this gives us a closed-form value for joint entropy.

Another way to get at this value is to use conditional entropy. For the case in which all information elements are independent, the joint entropy calculation is additive so that

$$H(X_1,X_2,\cdots X_C)=\Sigma_{i=1}^{C}H(X_i).$$

This is, in effect, an upper bound on joint entropy so that, in general,

$$H(X_1,X_2,\cdots X_C)\le\Sigma_{i=1}^{C}H(X_i).$$

However, if the conditional entropies can be calculated, joint entropy can be calculated directly as

$$H(X_1,X_2,\cdots X_C)=\Sigma_{i=1}^{C}H(X_i\,|\,X_{i-1},\cdots X_1).$$

Summing Up

The degree of uncertainty in a cluster depends on the information collection assets devoted to the cluster's critical elements of information and the extent to which collaboration among the cluster decision nodes is facilitated. We apply a probabilistic entropy model to represent the uncertainty associated with the critical information elements

needed within the cluster. The reports from sensors or other information-gathering sources are treated as estimates of the means of the distributions describing their uncertainty.

These estimates are transformed into awareness and knowledge and form the basis of decisionmaking. The metrics we have developed in this chapter quantify this process through the use of information entropy and the derivative knowledge metrics.

Information sharing among nodes ideally tends to lower information entropy (and hence increases knowledge) because of the reduction in variance and the buildup of correlations among the critical information elements. One of the key aspects of increased knowledge is increased understanding.

A key requisite for calculating cluster (and eventually network) knowledge is an acceptable method for combining knowledge gained from all critical information elements at a single headquarters, how that combination produces cluster knowledge, and how cluster knowledge combines to produce network knowledge. We have suggested several combining techniques, several of which require knowledge of a joint probability distribution. In many cases, the joint probability distribution of all critical information elements is not known and is difficult, if not impossible, to calculate empirically.

CHAPTER FOUR

The Effects of Collaboration

Networks provide an opportunity for participating decision entities to collaborate by sharing information as they form clusters. This is generally thought to be a good thing, as we have seen so far. By sharing, we experience synergistic effects that improve what we know (the completeness of our information) and how accurately we know it, as measured in terms of its bias and precision. In other words, collaboration improves both the quantity and quality of the information we need to take decisions. As compelling as this argument is, there is also a possible negative aspect of collaboration and information sharing: the expenditure of resources needed to deal with information overload and disconfirming evidence. The former can lead to processing delays and the latter to indecision as contradictory information is resolved. We treat these in more detail in Chapter Five. Here the focus is on the role of collaboration across a cluster in producing information that is complete and accurate.

Knowledge

As discussed earlier, information entropy appears to be a good surrogate for assessing the knowledge level within a cluster about a given critical information element. Until now, we have focused only on knowledge derived from the entropy associated with the probability distribution for the uncertain information elements. As noted earlier, the entropy function is always a function of the distribution variance,

and therefore our knowledge function is a function of precision only; that is, it measures the degree to which the observations of the critical information element are 'close together'. To assess the degree to which the networked headquarters affect decisionmaking, our measure of knowledge must also include the completeness and the bias of the estimates. We thus begin by examining the components of accuracy, namely, *bias* and *precision.*

Bias

Bias in an estimate is error introduced by systematic distortions. For example, suppose we were to conduct an experiment in which the temperatures of some substance had to be measured over time. If the thermometer we used were calibrated such that every reading was off by 1 degree Celsius, the result would be a set of biased measurements. Bias therefore is systemic, not random, error.

An unbiased estimator therefore is one for which $E[\mu]=\mu$. That is, the expected value of the estimate of the parameter, $\hat{\mu}$, is the true value of the parameter, μ. Thus, the bias in the estimate is the degree to which this is *not* true, or $b=|E[\hat{\mu}]-\mu|$.

Precision

The variation in estimates of the critical information elements can occur in a purely random way. For example, an observer may make an error in judgment such as reporting a tracked personnel carrier to be a tank. The variation may also be the result of uncontrollable environmental conditions, such as weather patterns, that cause sensor occlusions. In any event, random errors of this kind affect the precision of the estimates reported because they increase the variance of the distribution of the estimated information element. In general, precision is defined to be the degree to which estimates of the critical information element(s) are close together.[1] Bias and precision, there-

[1] This is a commonly accepted definition. Ayyub and McCuen (1997, p. 191) define precision as 'the ability of an estimator to provide repeated estimates that are very close together'. A similar definition can be found in Pecht (1995).

fore, are independent. That is, biased estimates may or may not be precise.

Precision and Entropy

In Appendix A, we describe the Rapid Planning Process by way of a simple operational example based on a local perceived force ratio.[2] For a given cluster i, at intermediate time steps j, we need only pursue the process as far as assessing the probability of each fixed pattern within the conceptual space of a local decisionmaker within the cluster. (These are the stored situations depicted in Figure 2.1.) The estimate produced is declared to be one of the stored situations, provided the estimate falls within the decisionmaker's comfort zone.

The joint probability density $f(\mathbf{x}_i(j))$, a multivariate normal distribution with covariance matrix $\boldsymbol{\Sigma}$, reflects the uncertainty associated with the critical information elements shared across cluster i at time step j, where $\mathbf{x}_i(j)=[x_{i,1}(j),x_{i,2}(j),\cdots,x_{i,C}(j)]$, the vector of perceived values of the critical information elements in the shared conceptual space, assuming all C elements are available to the cluster. This is the shaded area labelled 'current estimate' in Figure 2.1 (Chapter Two).

The mean of the current estimate, $f(\mathbf{x}_i(j))$, reflects the current best estimate, based on reports received from organic sources and information shared across the cluster, and the covariance, $\boldsymbol{\Sigma}$, reflects the *precision* of this estimate. The amount of information available in the joint (multivariate normal) probability density is measured in terms of the relative information entropy, $H(\mathbf{x}_i(j))=\log|\boldsymbol{\Sigma}|$. Both precision and information entropy are a function of the covariance.

[2] The perceived force ratio is calculated from the recognised picture, generally defined by a number of attributes (elements of information). A detailed description of both the recognized picture and the perceived force ratio can be found in Chapter 2 of Moffat (2002).

Estimating Local Knowledge

Local knowledge is a measure of understanding within cluster *i*.[3] As demonstrated earlier, there is an inverse relation between entropy and knowledge based on precision alone: As entropy increases, knowledge decreases. In general, we get the following knowledge metric for a joint distribution of the information elements, which is multivariate normal:[4]

$$K(\mathbf{x}_i(j)) = 1 - e^{-\left[\log|\boldsymbol{\Sigma}|_{\max} - H(\mathbf{x}_i(j))\right]}$$
$$= 1 - \frac{|\boldsymbol{\Sigma}|}{|\boldsymbol{\Sigma}|_{\max}}$$

where $\log|\boldsymbol{\Sigma}|_{\max}$ is the maximum relevant information entropy and $|\boldsymbol{\Sigma}|_{\max}$ is the determinant of the corresponding covariance matrix. $f(\mathbf{x}_i(j))$ reflects the level of understanding within the cluster based on precision alone.

Precision and Knowledge in the Logistics Example

To illustrate, we continue with the logistics example from Chapter Two. In Figure 2.3a, when there is no collaboration among the nodes, each is monitoring its requirement for fuel and providing estimates to the single master decision node. This configuration produces a single cluster comprising only the master node. When the nodes are collaborating, as in Figure 2.3b, information is shared between the two nodes, and therefore we take them to be dependent. The network in this case is a single cluster consisting of the two demand nodes and the arbiter node.

---

[3] By *understanding*, we mean the ability of humans to draw inferences about the possible consequences of a situation. Clearly, knowledge enhances this ability and therefore can be considered a measure of understanding.

[4] Actually, the exact entropy value for the bivariate normal case is $H(x,y) = \log|(2\pi e)^2\Sigma|$. However, because we are concerned about the ratio of entropies, we use the simpler version.

For simplicity, we start by assuming that the fuel requirement is normally distributed.[5] Consequently, we let a_1 be the information element 'fuel demand for node 1' and a_2 be 'fuel demand at node 2'. The fuel supply at the master node is assumed to be known with certainty; that is, the master node is self-aware in both cases. Consequently, the critical information element set is $\mathrm{A}=\{a_1,a_2\}$, and the value of each is depicted as $\mathbf{x}=[x_1,x_2]^T$. In each case, the distribution of uncertainty about the information element is normal with mean $\boldsymbol{\mu}=[\mu_1,\mu_2]^T$ and covariance

$$\Sigma=\begin{bmatrix}\sigma_1^2 & \rho_{1,2}\sigma_1\sigma_2\\ \rho_{1,2}\sigma_1\sigma_2 & \sigma_2^2\end{bmatrix}.$$

The fuel levels at each of the demanding nodes may be correlated, and the effects of collaboration are therefore dependent on the correlation coefficient, $-1\le\rho_{1,2}\le1$.[6] If, as in Figure 2.3a, the nodes are not collaborating, the 'network' consists of the master node and the two isolated demand nodes, and we model the lack of collaboration by setting $\rho_{1,2}=0$. That is, we assume that each headquarters is acting independently, and therefore all their demands for fuel are independent. The off-diagonal elements in the covariance matrix then are 0. This is clearly the simplest case to analyse because the implications of 'no collaboration' are clear in that it produces uncorrelated fuel levels. In cases like Figure 2.3b, where collaboration between the nodes takes place and $\rho_{1,2}\neq0$, collaboration can be shown to have a salutary effect.

In general, total cluster information entropy is

$$H(\mathbf{x})=\log|\Sigma|=\log\left[\sigma_1^2\sigma_2^2\left(1-\rho_{1,2}^2\right)\right].$$

---

[5] This is valid only when the mean demand is large and the variance is suitably small.

[6] Although the fuel is received from the same source, the demand may be generated independently and therefore may be uncorrelated.

In the non-collaboration case, this reduces to $H(\mathbf{x}) = \log(\sigma_1^2\sigma_2^2)$. Because entropy measures the degree of uncertainty in probability distributions, small values of $H(\mathbf{x})$ are desirable. Regardless of the values of the variances σ_1^2 and σ_2^2, this will occur when $|\rho_{1,2}|$ is near 1.0. Conversely, maximum entropy and, therefore, maximum uncertainty occur for $\rho_{1,2} = 0$. The change in entropy from the non-collaboration case and the collaboration case then is

$$H_a(\mathbf{x}) - H_b(\mathbf{x}) = -\log\left(1 - \rho_{1,2}^2\right).$$

As before, a value of $|\rho_{1,2}|$ close to 1.0 maximises this quantity.

To convert entropy to knowledge, we first need to establish a maximum entropy value,[7] which is equivalent to establishing a maximum variance or determinant of the covariance matrix as developed in Chapter Three. Because the maximum covariance occurs when the random variables are uncorrelated, we have

$$H_{\max}(a_1, a_2) = \log\left(\sigma_{1,\max}^2 \sigma_{2,\max}^2\right).$$

We can now develop the knowledge metric to measure the current level of understanding of the fuel demand for both the collaboration (*a*) and non-collaboration (*b*) cases. We have for the non-collaboration case that

$$\begin{aligned} K_a(\mathbf{x}) &= 1 - e^{-\left[\log \sigma_{1,\max}^2 \sigma_{2,\max}^2 - \log \sigma_1^2 \sigma_2^2\right]} \\ &= 1 - \frac{\sigma_1^2 \sigma_2^2}{\sigma_{1,\max}^2 \sigma_{2,\max}^2}. \end{aligned}$$

[7] This is necessary because entropy for continuous random variables (referred to as *differential entropy*; see Appendix B) is always unbounded.

For the collaboration case, we have

$$K_b(\mathbf{x}) = 1 - e^{-\left\{\log\left[\sigma_{1,\max}^2 \sigma_{2,\max}^2\right] - \log\left[\sigma_1^2 \sigma_2^2 \left(1-\rho_{1,2}^2\right)\right]\right\}}$$
$$= 1 - \frac{\sigma_1^2 \sigma_2^2 \left(1-\rho_{1,2}^2\right)}{\sigma_{1,\max}^2 \sigma_{2,\max}^2}.$$

The benefits of collaboration therefore are measured as the difference between the two, or the increase in understanding represented by

$$\Delta K(\mathbf{x}) = K_b(\mathbf{x}) - K_a(\mathbf{x}) = \frac{\rho_{1,2}^2 \sigma_1^2 \sigma_2^2}{\sigma_{1,\max}^2 \sigma_{2,\max}^2}.$$

Here the buildup in correlation between the information elements a_1 and a_2 causes the increase in knowledge. This relates to our commonsense understanding of an increase in knowledge of our surroundings because we know how one thing relates to another.

Accuracy

Accuracy is the degree to which the estimates of the critical information elements are close to ground truth. Collaboration across a cluster affects the accuracy of the estimates of the information elements—and hence the degree to which fixed patterns in the shared conceptual space are indeed ground truth. The concept of accuracy comprises both precision and bias: The smaller the bias, the closer the estimate is to ground truth, and the more precise the estimates (i.e., the more closely they are grouped), the more confident we are in the estimate.

We generally take ground truth to be the ideal distribution mean for the information and measure the bias of the estimates in terms of the distance from the ground-truth value. This assumes, of course, that we know ground truth, which is always the case in models and simulations aimed at assessing and comparing alternative C4ISR (command, control, communications, computers, intelli-

gence, surveillance, and reconnaissance) systems, alternative network constructs, or alternative operating procedures. Support to actual operations in which ground truth is not known requires an assessment of the *consistency* of the information reported. In some cases, we may instead take as our point of comparison the estimates coming from the higher-level planning process. For our fuel logistics illustration, we can, for example, compare the perception of the fuel demand from each of the two nodes, with the assessment made from the top-down planning assumptions.

In general, if a is an information element whose value, x, is unknown with probability distribution $f(x)$ and mean μ representing ground truth, then the bias associated with the estimate of the mean is $b = |E(\hat{\mu}) - \mu|$, where $\hat{\mu}$ is the estimate of the mean based on one or more reports on the value of a. Because accuracy consists of both bias and precision, we need a metric that combines both. One such metric is the mean square error (MSE), defined as $E[(\hat{\mu}-\mu)^2]$. It can be shown that $E[(\hat{\mu}-\mu)^2] = b^2 + \sigma^2$, where σ^2 is the variance of $\hat{\mu}$.[8] This metric is extremely useful because it includes both accuracy in the total and precision as a component. In estimating ground truth, the bias accounts for nonrandom errors and the precision accounts for random errors.

To illustrate this, in our CEC cluster example, suppose we want to estimate the location of an enemy cruise missile based on several sequentially arriving reports from the collaborating team. Each report is processed in turn using Bayesian updating to refine the location estimate.[9] In this case, we need an estimate for both the x-coordinate and the y-coordinate. The bivariate normal distribution is used to represent the uncertainty associated with the random location vector, $\mathbf{x} = [x, y]^T$. The estimator in this case is a Bayesian estimator of the form:

---

[8] See, for example, Cover and Thomas (1991).

[9] Later in this chapter, we suggest maintaining the incoming reports and variance estimates until a decision is imminent. If we perform the updates sequentially at that time, we can account for the age of the reports—essentially discounting older reports.

$$\boldsymbol{\mu}_{t+1} = \boldsymbol{\Sigma}_t \left(\boldsymbol{\Sigma}_t + \hat{\boldsymbol{\Sigma}} \right)^{-1} \hat{\boldsymbol{\mu}} + \hat{\boldsymbol{\Sigma}} \left(\boldsymbol{\Sigma}_t + \hat{\boldsymbol{\Sigma}} \right)^{-1} \boldsymbol{\mu}_t \tag{4.1}$$

and

$$\boldsymbol{\Sigma}_{t+1} = \boldsymbol{\Sigma}_t \left(\boldsymbol{\Sigma}_t + \hat{\boldsymbol{\Sigma}} \right)^{-1} \hat{\boldsymbol{\Sigma}}. \tag{4.2}$$

In this formulation, $\hat{\boldsymbol{\mu}}$ is the collaborating team's current estimate of $\boldsymbol{\mu} = [\mu_x, \mu_y]^T$. The instrument (sensor, source, process) used to obtain the estimate has an associated error, which we record as $\hat{\boldsymbol{\Sigma}}$, an estimate of the variance. This may be acquired from the target location error associated with sensor or source and existing environmental conditions prevailing when the measurement was made.[10] This matrix serves as a weight. For large $\hat{\boldsymbol{\Sigma}}$, the expression $\boldsymbol{\Sigma}_t(\boldsymbol{\Sigma}_t + \hat{\boldsymbol{\Sigma}})^{-1}$ is very small (close to the zero matrix) and $\hat{\boldsymbol{\Sigma}}(\boldsymbol{\Sigma}_t + \hat{\boldsymbol{\Sigma}})^{-1}$ is approximately the identity. Therefore, the current report has little influence on $\boldsymbol{\mu}_{t+1}$. The reverse is true for small $\hat{\boldsymbol{\Sigma}}$. The initiating estimates, $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$, are obtained from the IPB process or are estimates prevailing at the last decision point.

The task now is to assess just how accurate the estimate is. If we are conducting a controlled experiment, such that the true location of the unit is known, then, as mentioned above, we can take advantage of the fact and calculate the bias in the estimate. In this case, the bias is the Euclidean distance between the Bayesian estimate and the ground-truth location of the unit, or

$$b = \sqrt{\left(\mu_{t+1,x} - \mu_x\right)^2 + \left(\mu_{t+1,y} - \mu_y\right)^2}\,.$$

[10] It is also possible to use the sample mean of several reports as an estimate of the latest of several reports, the 'best' report, etc. Each will require an accompanying estimate of the variance.

For our purposes, we can take this estimate to be the true location of the enemy cruise missile at a specific moment in time. By analogy with the MSE, the accuracy of the estimate is defined as

$$D(\mathbf{x}) = b^2 + \left|\boldsymbol{\Sigma}_{t+1}\right|.$$

Accuracy in the Logistics Example

Recall that the amount of fuel required at each node is $\mathbf{x} = [x_1, x_2]^T$. We assume the $\mathbf{x}$ is bivariate normal with mean $\boldsymbol{\mu} = [\mu_1, \mu_2]^T$. Reports on projected fuel requirements are processed sequentially to create a current estimate of future requirements for both nodes. As in the location estimate discussed above, the estimator is Bayesian and the bias is the Euclidean distance between the estimates and ground truth. However, unlike the location example above, the error associated with each report is generated from two sources. In the first case (no collaboration), the errors are independent, and in the second case (collaboration) they are not. We also assume that a report is received from both nodes near-simultaneously.

The estimate covariance matrices depend on the model selected and the update methodology. For example, in the no-collaboration case (Figure 2.3a), the sample covariance matrix is

$$\boldsymbol{\Sigma}_a = \begin{bmatrix} \hat{\sigma}_1^2 & 0 \\ 0 & \hat{\sigma}_2^2 \end{bmatrix}.$$

When they are collaborating, as in Figure 2.3b, the sample covariance matrix is

$$\hat{\boldsymbol{\Sigma}}_b = \begin{bmatrix} \hat{\sigma}_1^2 & \hat{\rho}_{1,2}\hat{\sigma}_1\hat{\sigma}_2 \\ \hat{\rho}_{1,2}\hat{\sigma}_1\hat{\sigma}_2 & \hat{\sigma}_2^2 \end{bmatrix}.$$

Generating these estimates may be problematic. The sources of the reports are generally the headquarters themselves, so the errors are associated with both the assessment of fuel on hand and future requirements. Standards may exist for predicting fuel requirements that vary with a unit's posture. In any event, it will be necessary to provide error estimates in both cases.

The bias is the Euclidean distance, as previously discussed, so that

$$b=\sqrt{\left(\mu_{t+1,1}-\mu_1\right)^2+\left(\mu_{t+1,2}-\mu_2\right)^2}\,.$$

At any time, the estimates of the covariance for both cases are

$$\boldsymbol{\Sigma}_{t+1,a}=\begin{bmatrix}\sigma^2_{t+1,1} & 0\\ 0 & \sigma^2_{t+1,2}\end{bmatrix}$$

and

$$\boldsymbol{\Sigma}_{t+1,b}=\begin{bmatrix}\sigma^2_{t+1,1} & \rho_{t+1}\sigma_{t+1,1}\sigma_{t+12}\\ \rho_{t+1}\sigma_{t+1,1}\sigma_{t+1,2} & \sigma^2_{t+1,2}\end{bmatrix},$$

and therefore the accuracy metrics for the two cases are (iteration subscripts omitted)

$$D_a(\mathbf{x})=b^2+\sigma_1^2\sigma_2^2 \text{ and } D_b(\mathbf{x})=b^2+\sigma_1^2\sigma_2^2-\rho^2\sigma_1^2\sigma_2^2.$$

Consequently, the increase in accuracy in the collaboration case is $\rho^2\sigma_1^2\sigma_2^2$. Again, this quantity is maximised when $|\rho|$ is close to 1.0. The task now is to measure these effects on cluster and network knowledge.

The Effects of Bias, Precision, and Accuracy on Knowledge

One way to account for bias, precision, and, hence, accuracy in the knowledge function is to replace the distribution variance with the MSE, or the accuracy measure, $D(\mathbf{x})$, in the knowledge function. Doing so has the effect of increasing the variance to account for bias. The MSE is bounded from below by the variance, so when the bias is 0, the MSE is just the variance. In the general case, we express knowledge as

$$K_M(X) = 1 - e^{-[H_{\max,M}(X) - H_M(X)]},$$

where $K_M(X)$ is the knowledge function with the variance replaced by the MSE. To do this, we calculate the maximum and current entropies in the usual way and then replace the variance (or more generally, the covariance) with the MSE.

For the multivariate normal case, for example, we get a modified knowledge function of the form:

$$K_M(\mathbf{x}) = 1 - \frac{b^2 + |\Sigma|}{\left(b^2 + |\Sigma|\right)_{\max}}.$$

The 'maximum' MSE is a combination of the maximum bias and the maximum precision and represents the maximum in inaccuracy. Because bias and precision are independent, the maximum occurs when both are maximised, or $(b^2 + |\Sigma|)_{\max} = b^2_{\max} + |\Sigma|_{\max}$. Like the variance, a suitable upper bound for bias can be found by searching for the largest possible measurement error the sensors or sources might produce.

We can apply this to the simple logistics problem. For the non-collaboration case, we get

$$K_{M_a}(\mathbf{x}) = 1 - \frac{b^2 + \sigma_1^2\sigma_2^2}{b^2_{\max} + \sigma^2_{1,\max}\sigma^2_{2,\max}}.$$

For the collaboration case, we have

$$K_{M_b}(\mathbf{x}) = 1 - \frac{b^2 + \sigma_1^2\sigma_2^2 - \rho^2\sigma_1^2\sigma_2^2}{b_{\max}^2 + \sigma_{1,\max}^2\sigma_{2,\max}^2}.$$

The maximum MSE is the same for both cases, given that the covariance is maximised when $\rho = 0$ and the variances are fixed. The difference between the two now reflects the effects of collaboration on knowledge as measured by precision, accuracy, and bias and is calculated to be

$$K_{M_a}(\mathbf{x}) - K_{M_b}(\mathbf{x}) = \frac{\rho^2\sigma_1^2\sigma_2^2}{b_{\max}^2 + \sigma_{1,\max}^2\sigma_{2,\max}^2}.$$

As expected, this quantity is diminished over the previously calculated values that considered precision only. However, if the estimate is unbiased ($b = 0$), the results are the same. Also, in the rare case that an estimate is reported as ground truth (no variance), bias is still possible so that there is no improvement in knowledge from the non-collaboration to the collaboration case.

We next discuss the contribution to information sharing of the completeness of the information available to take a decision. The combining of precision, accuracy, bias, and completeness then will measure the overall contribution of collaboration across the cluster to knowledge and thus to improved local decisionmaking.

Completeness

For any cluster i, we have defined the complete data set at time t as the set $\mathbf{x}_i(j) = [x_{i,1}(j), x_{i,2}(j), \cdots, x_{i,C}(j)]$. The set consists of a maximum of C elements of critical information; however, only a subset consisting of $n \leq C$ out of C elements might be available at time t. If waiting for additional reports is not possible, a decisionmaker would be required to take a decision without the benefit of complete information. Depending on his experience and other contextual informa-

tion, he may be able to infer some likely less reliable value for the missing information. For now, we assume that if the value of an information element is missing, the value of completeness for cluster i is

$$X_{i,t}(n)=\left[\frac{n}{C}\right]^{\xi},$$

where ξ is a 'shaping' factor that reflects the decisionmaker's aversion to risk because the selection of the appropriate value depends on the consequences, as perceived by the decisionmaker, of being forced to take a decision with incomplete information. For values of $\xi<1$, the curve is concaved downwards, thus reflecting a high aversion to risk; for $\xi>1$, it is concaved upwards, reflecting little aversion to risk; and for $\xi=1$, it is a straight line, reflecting the decisionmaker's equivocation concerning risk. The ultimate impact of this lack of completeness is the uncertainty of the decisionmaker's perception of where he is in the conceptual space, as depicted in Figure 2.1. The selection of the appropriate values depends on the consequences associated with being forced to take a decision with incomplete information.

With the addition of completeness, we are now ready to combine the measures of collaboration, namely accuracy (i.e., precision and bias) and completeness, to produce a single knowledge metric that can be subsequently combined with the measures of complexity discussed in Chapter Five. But before we develop the combined collaboration metric, we must first address another measure of information quality: its currency. It is generally assumed that more recent, or *fresher*, information is desirable over older information. This supposition is certainly true in the modern battlespace, where events change rapidly.

Information 'Ageing'

The information gathered by a cluster consists of reports concerning one or more of the critical information elements shared across the cluster, which are necessary to take a decision.[11] These reports are used to update the joint probability distribution of uncertainty concerning these information elements. If the reports are old, we assume that their contribution to reducing uncertainty is less than if they are fresh. In addition, if resources within a cluster are such that reports arriving are not processed in a reasonable period of time, they will age in a queue with the same effect.

Freshness is a consideration that is separate from timeliness. Freshness is concerned with how old the information is and, as such, is generally context free. Timeliness, however, deals with when the information is needed and, as such, is situation dependent. Both timeliness and freshness are functions of the time volatility of information, i.e., the rate at which information is likely to change over time. For example, consider assessing the location of a missile versus the location of a mountain. Information about the location of a mountain is considered time resilient, and therefore freshness and timeliness are essentially equivalent. However, we take the position here that the older the information is, the lower its quality.

Precision, bias, and, hence, accuracy depend on the estimator selected (a Bayesian estimator in this case) to estimate fixed patterns of ground truth shared across a cluster. They are also dependent on the joint probability density function that reflects the uncertainty in our knowledge. Consequently, what is needed is a methodology that allows us to incorporate the age of the reports in our updating process.

Time Lapse

For each critical information element, $a_{i,j}$, shared across cluster i at time period j, we record the time that its estimated value, $x_{i,j}$, was

[11] In Chapter Five, we address the issue of unneeded information and its contribution to 'information overload'.

last reported and shared across cluster i, $t_{i,j}$. If a decision within cluster i is to take place at time t_i, then $F_{i,j} = t_i - t_{i,j}$ is a measure of the freshness of information element at the time it is used. We can further express the importance of freshness by an exponent so that we get

$$F_{i,j} = \left(t_i - t_{i,j}\right)^{\eta},$$

where the parameter $\eta > 0$ reflects the degree to which freshness of a report concerning information element a_i at time period j is an important consideration in taking a decision, i.e., the time volatility of the information. For example, the freshness of information concerning the location of the Baath Party's headquarters in An Nasiriyah is not as critical as a report on the location of Fedayeen Saddam forces in the city.

To be consistent with other metrics, we choose to normalise $F_{i,j}$ as follows:

$$\Phi_{i,j} = \left[\frac{t_i - t_{i,j}}{t_i - t_0}\right]^{\eta},$$

where t_0 is the time at which the data collection begins in cluster i for this decision. In the case of the Baath Party headquarters, a value of $\eta \geq 1$ would be appropriate. In the case of the Fedayeen, we would place considerable importance on fresh information and therefore assign a value of $\eta < 1$.

Updating

Within the time required to take a decision within cluster i, several reports from sensors and sources of the estimated value of $a_{i,j}$ are likely to be produced—each with time-lapse estimate $\Phi_{i,j}$, calculated as above. By updating the value of the information element over time, we can also account for the age of the data reported. In this way, we directly affect the information and therefore its knowledge

function. As mentioned earlier, we have elected to update our estimates using Bayesian updating.

We assume that at some time t, U sequentially arriving reports concerning information element a, $\{(\hat{\mu}_1,\hat{\sigma}_1^2),(\hat{\mu}_2,\hat{\sigma}_2^2)\cdots,(\hat{\mu}_U,\hat{\sigma}_U^2)\}$ are to be combined to support a decision to be made at time t.[12] The pairs, $(\hat{\mu}_k,\hat{\sigma}_k^2)$ are the kth sequential estimate of the mean and variance of the distribution describing the uncertainty in the value of the information element a. The scalar versions of equations (4.1) and (4.2), uncorrected for freshness, are

$$\mu_{k+1} = \frac{\sigma_k^2\hat{\mu}_k + \hat{\sigma}_k^2\mu_k}{\sigma_k^2 + \hat{\sigma}_k^2}$$

and

$$\sigma_{k+1}^2 = \frac{\sigma_k^2\hat{\sigma}_k^2}{\sigma_k^2 + \hat{\sigma}_k^2}.$$

The pair (μ_0,σ_0^2) are the estimates existing at time t_0. This could be the IPB estimate, or it could be the estimate at the last decision. Next, we modify the estimates to account for the freshness of the reports.

We assume that the effect of ageing makes the estimate less certain. Ageing therefore is a function of the estimated variance coupled with the normalised freshness factor, Φ_k. For the more recent reports, Φ_k is small, and for older reports, it is large. In any case, $0 \le \Phi_k \le 1$, which suggests a net present value model for measuring the effect of Φ_k on the variance of the estimate; that is, we replace the variance with $(1+\Phi_k)\hat{\sigma}_k^2$.[13] This yields the following modified Bayesian update

[12] We drop the cluster and time period subscripts for clarity. It should be clear that the information element is required at cluster i and that the time period at which the combining takes place is j.

[13] The net present value, P, of a principal amount, A, compounded over n time periods is $P = A(1+i)^n$, where i is the rate of return. The argument for an analogous approach in this case is that freshness can be thought of as the rate at which the variance increases.

formulas for producing the current estimate for the value of the information element *a*:

$$u_{k+1} = \frac{\sigma_k^2 \hat{\mu}_k + (1+\Phi_k)\hat{\sigma}_k^2 \mu_k}{\sigma_k^2 + (1+\Phi_k)\hat{\sigma}_k^2}$$

and

$$\sigma_{k+1}^2 = \frac{\sigma_k^2 (1+\Phi_k)\hat{\sigma}_k^2}{\sigma_k^2 + (1+\Phi_k)\hat{\sigma}_k^2}.$$

In the best case, when the freshness factor is 0 (the report arrives at decision time), there is no effect on the reported variance. In the worst case, when the report dates to the beginning of collection for the current decision, the reported variance is doubled.

The final estimate, μ_U, calculated in this way is taken to be the estimate of the true mean of the distribution with bias, and variance estimate, σ_U^2. The updated density function is therefore $f(x:\mu_U,\sigma_U^2)$. From this we can calculate a current, updated knowledge estimate, $K_M(x)$.[14]

Measuring the Overall Effect of Cluster Collaboration

Finally, we combine the currency adjusted precision and accuracy knowledge function with completeness to arrive at a single metric to assess the effects of collaboration across the cluster. The ideal case is when we have full completeness, i.e., $X_t(n) = X_t(C) = 1$, and the knowledge shared across the cluster is fully accurate, i.e., $K_M(\mathbf{x}) = 1$, for the multivariate case. In this case, collaboration is able to provide complete information, and its accuracy provides the local

[14] In the special case of a multivariate normal distribution of uncertainty across the information elements, this effect can be put in place by adjusting the initial values of the 'observation noise' and 'system noise' in the DLM (as discussed in West and Harrison, 1997).

decisionmaker with perfect knowledge or situational awareness. Unfortunately, this ideal is seldom, if ever, achieved. Consequently, we require a construct that gauges the degree to which accuracy, as calculated here, and completeness contribute to knowledge.

The knowledge function, $K_M(\mathbf{x})$, is derived by replacing the variance in the entropy calculation with the MSE, thus allowing us to account for both precision and bias. It is logical that we proceed in the same way with completeness; that is, we replace the MSE with a function of the MSE and completeness. In general, when $X_t(n)$ is small, (i.e., when there exists estimates for only a small fraction of the required number of information elements), the knowledge function should also be small, all things being equal, because this means that the aggregate accuracy of the estimates is based on only a few elements of information. One way to reflect this behaviour is to replace the MSE in the entropy calculation with

$$\frac{b^2+\sigma^2}{X_t(n)}.$$[15]

This calculation has the desirable property that when $X_t(n) \to 1.0$, the ratio is just the MSE, and that when $X_t(n) \to 0$, it increases without bound. This indeed reflects the fact that if we have no information, we have no knowledge and thus the bias and variance estimates are irrelevant. However, it is not practical to use this calculation as a lower bound, since it will drive the ratio

$$\frac{b^2+\sigma^2}{X_t(n)} \Big/ \frac{\left(b^2+\sigma^2\right)_{\max}}{\left[X_t(n)\right]_{\max}}$$

[15] Although we illustrate the discussion with the univariate case, this applies equally to the multivariate case.

to increase without bound for all values of n. To avoid this, we arbitrarily select $n=1$ to be the worse case with $X_t(1)=C^{-\xi}$.[16] Consequently, the upper bound on the resultant entropy calculation is

$$\frac{b_{\max}^2+\sigma_{\max}^2}{C^{-\xi}}=C^{\xi}\left(b_{\max}^2+\sigma_{\max}^2\right).$$

This has the effect of increasing the maximum MSE when the requisite number of information elements is large. Note that if $C=1$, there is no effect on the current entropy calculation or on the maximum entropy. If we let $K_\kappa(x)$ be the knowledge within the cluster based on accuracy and completeness, then

$$K_\kappa(x)=1-e^{-\left[H_{\kappa,\max}(x)-H_\kappa(x)\right]},$$

where $H_{\kappa,\max}(x)$ is the entropy calculated with the maximum variance replaced with $C^{\xi}(b_{\max}^2+\sigma_{\max}^2)$ and $H_\kappa(x)$ is the current entropy calculated with the current variance replaced with

$$\frac{b^2+\sigma^2}{X_t(n)}=\left[\frac{C}{n}\right]^{\xi}\left(b^2+\sigma^2\right).$$

Applying this to the normal case, we get the knowledge gained when completeness is accounted for as

$$K_\kappa(x)=1-\frac{b^2+\sigma^2}{n^{\xi}\left(b_{\max}^2+\sigma_{\max}^2\right)}.$$

Knowledge increases when the values of more of the requisite information elements have been reported and is maximised when $n=C$. This simple formulation is intuitively satisfying because we would expect that for all the precision and accuracy, unless information on all the information elements is present, our knowledge will be deficient. This scales naturally to the multivariate normal case as

[16] We discuss the special case of $C=1$ later.

$$K_{\kappa}(\mathbf{x}) = 1 - \frac{b^2 + |\Sigma|}{n^{\xi}\left(b_{\max}^2 + |\Sigma|_{\max}\right)}.$$

Up to this point, we have captured the effects of collaboration among decision nodes within a cluster on knowledge. The measured effects of information sharing through collaboration are accuracy and completeness. For the most part, these effects are dynamical, since they vary with the quality and quantity of reports received and processed over time. Missing from this analysis so far has been an assessment of the systemic effects of the network architecture, effects that are more static. In the next chapter, we take up such measures of network complexity and combine them with the collaborative effects to arrive at a single measure of network performance and its effect on decisionmaking.

CHAPTER FIVE
The Effects of Complexity

In the previous chapter, we were concerned about measuring the effects of collaborative decisionmaking among the decision nodes within a cluster. Although the ability to collect, process, and share information is dependent on the structure of the supporting network, we focused our assessment on the dynamics of operations: the effects of processed and shared information over time. In this chapter, we focus on the network itself and its ability to enable efficient and effective information flow. Our measure is complexity, and we examine both the detrimental effects of overly complex networks and the salutary effects of complexity.[1]

Complex Networks

All networks are complex to a greater or lesser degree, including military command and control systems operating in a network-centric environment. The challenge is to understand the nature of complexity, what its effects are, and how to quantify them. All three tasks have been attacked in the past (we briefly summarise a history below); however, a satisfactory resolution is still elusive. One thing is certain,

[1] Much of the discussion on complexity in this chapter is taken from an unpublished RAND report: W. Perry, F. Bowden, J. Bracken, R. Button, J. McEver, and T. Sullivan, *Advanced Metrics for Network-Centric Naval Operations*, December 2002 (J. McEver contributed the work on complex systems in the referenced report.)

though: There are both good and bad effects of complexity. For this reason, we have adopted Murray Gell-Mann's more neutral term *plecticity* to describe the effects of the network infrastructure on military operations. This characterisation avoids the negative aspects of the term 'complexity'.[2]

What Is Complexity?

Complex networks (such as the World Wide Web, which operates on another complex network, the Internet), have been studied for years in attempts to understand their structure and properties. The science of complex adaptive systems, too, has evolved in less than two decades as an interdisciplinary attempt to understand how components, when tied together in certain ways, yield systems with capabilities different from those of their constituent components taken separately.[3] Still, although general agreement exists on what, broadly, complexity is, there are no agreed-on definitions of complexity, much less quantitative measures of complexity in networks.

For decades, researchers have recognised that as systems grow and become more complicated, their behaviour departs substantially from that of the system's components (Anderson, 1972). In 1965, Kolmogorov proposed a useful definition of complexity: 'The complexity of an object is the shortest binary computer program that describes the object' (Kolmogorov, 1965). It can be shown that, defined in this way, complexity is approximately equivalent to Shannon entropy, a well-defined mathematical construct described earlier (Shannon, 1948). Shannon entropy, as a surrogate for complexity, is used in medical research to assess the complexity of biological systems. Other definitions, similar in spirit to the Kolmogorov com-

[2] Gell-Mann (1995/1996) argues that the study of complex adaptive systems is better referred to as plectics, because it is 'a broad, transdisciplinary subject covering many aspects of simplicity and complexity as well as the properties of complex adaptive systems, including composite complex adaptive systems consisting of many adaptive agents'. Gell-Mann derives the word 'plectics' from the Greek work *plektos*, which can refer to both simplicity and complexity. Invocation of the word plectics allows for the study of entanglement or the lack thereof.

[3] See, for example, Moffat (2003).

plexity, have been proposed, including Gell-Mann's *effective complexity*, defined as the length of the description of the regularities, or the 'grammar' of a system, and Bennett's logical depth, which defines complexity as the processing time theoretically required for a computer to go from the description of a system to the ability to duplicate the system itself (Gell-Mann, 1995).

In addition to these attempts at defining complexity, some quantitative definitions of complexity aimed at calculating the complexity of specific systems have been proposed. Again, no consensus definitions have emerged from the literature, which has the flavour of a spirited debate among many camps. Wolpert and Macready (1997) propose a quantification of how the spatio-temporal patterns of different scales of a system differ from one another ('self-dissimilarity') as a signature of system complexity. Sporns and Tononi (2002) describe a method they and Edelman developed to measure the complexity of the brain by relating functional segregation and integration measures. Solé and Luque (2002) discuss and refine a proposed stochastic-based complexity measure of nonlinear physical systems, based on the system entropy, the number of states to which the system has access, and a measure of the interaction between the components of the system. Other quantifications of complexity exist as well, and ultimately we too present a complexity metric in this work, specifically for a decision network, such as that proposed above, that can be applied to evaluate alternative network clustering structures.

Even though this literature has yielded useful insight into the problem of defining network complexity and understanding what features of a complex network result in the effectiveness and adaptability properties we desire, direct application of these complexity definitions has proven difficult. This is particularly true in the context of a command and control network in which the network components themselves are complex and adaptive and, specifically, do not have simple rules for how they interact with their network neighbours (as is assumed in models of complex physical systems). Instead, by combining insight from physics, medicine, and neural network approaches to complexity measurement with an understanding of network topology and desired decision network features, we can

move towards defining a metric for evaluating various network clustering possibilities.

Plecticity

In this context, *plecticity* refers to the ability of a connected set of actors to act synergistically via the connectivity between them. This measure is, in effect, the value added to the capability of the system by the entanglements between the system's nodes (decision nodes in this work). It is intended to take into account the fact that there may be constraints on how nodes can constructively connect to other nodes, because of either technical or procedural limitations. That is, a node's connectivity can add costs as well as benefits to network performance. Thus, networks can gain value both from the entanglements that are present and from those that are not. A measure of plecticity should account for the value of the nodes' ability to glean information from throughout the network to fulfil its particular functions, include a means for measuring the value of network redundancy, and reflect a cost to network effectiveness if nodes are overwhelmed.

Command and control networks that do well with regard to these measure attributes should be able to more readily enable the acquisition of timely information and facilitate a decisionmaker's more effective use of information resources gleaned from the network for the performance of mission functions.

Accessing Information

A decision network must provide the decisionmakers in a cluster the ability to gain easy access to information needed to support decisionmaking. Whether the information is 'pushed', as from sensors and sources, or 'pulled', as with queries over the Internet, the degree to which the information is accessible is an important measure of a network's effectiveness. Because accessibility is closely related to the completeness of information, we begin the development of the acces-

sibility metric from the completeness metric developed in the previous chapter.

The metric developed earlier for the completeness of the information set shared across the cluster is simply a ratio of counts: [available required information elements] to [total required information elements]. Therefore, no attempt was made to assess the degree to which we can really *expect* to receive the information element, i.e., the degree to which the network allows the cluster to access information in the network, or *information accessibility.* A metric that does this is the ratio of [the aggregate expected degree of critical information access] to [the total number of information elements across the network]. Such a metric accounts for the uncertainties associated with retrieving needed information. For our CEC cluster example, in maintaining an enemy missile track, the 'distance' required information must travel from source to destination might be used to assess the strength of the connectivity between the source and the destination for a given information element.[4] If we calculate the *connectivity,* k_l, for information element a_l in such a way that $0 \leq k_l \leq 1$, we arrive at a connectivity value $k \leq n$, with the equality holding only when the distance is negligible and the connectivity is robust. As before, n is the number of critical information elements for which at least one report has been made available. In this case, $k = f(k_l)$.[5] Although not technically a probability, connectivity calculated in this way does reflect the uncertainties associated with moving information through a network.

Another way to look at it is in terms of transmission *costs.* Replacing the binary accounting for information elements as was done in the completeness score, with a connectivity score based on a distance function of this sort, recognises the cost imposed by the path the information must take through the network to arrive at the cluster requiring it. That is, if, for a given network configuration, a speci-

[4] *Distance* in this context refers not only to the physical separation between source and destination, but may also include other factors such as the time required to move information.

[5] It is understood that the information element is critical to node i at time t. However, for ease of exposition, we omit these two subscripts.

fied type of information follows an 'expensive' path in getting from its source to the cluster requiring it, that source's contribution to the supply of information to the cluster takes a value lower than one that is less expensive. Consequently, the accessibility is diminished.

Distance and Connectivity

The distance function can be something as simple as the number of links in the path from source to sink. A more complicated function might take into account the individual capabilities of each link and node in the path. Because both nodes and links comprise a path's length, the more realistic assessment considers both. For now, we defer the mathematical construct of the distance function and focus on its use in constructing a connectivity metric. For any cluster information element, a_l, we are interested in the shortest path from source node to destination node, $d_l \geq 1$, however calculated.[6] The quantity, d_l, represents the expense incurred by moving information element a_l from source to destination. This value is now used to calculate the connectivity value, k_l, for information element a_l as follows:

$$k_l = \frac{1}{d_l^{\omega_l}},$$

where $\omega_l \geq 1$ is the rate at which k_l varies with changing values of the distance function by reflecting the importance of the distance d_l. To adequately determine a suitable value for ω_l, it is necessary to assess the relative importance of obtaining reports on information element a_l. Given that a costless direct connection between two nodes results from a distance cost score of $d_l = 1$, a strong connectivity score of $k_l = 1$ results. As the distance cost increases, the connectivity value approaches 0. If no path exists between any source of information element x_l and its destination, then $d_l \to \infty$ and $k_l = 0$.

[6] The restriction that the path distances always exceed 1.0 accounts for the fact that, for connectivity to exist at all, at least one link must exist between source and destination. The case in which no links exist implies an infinitely long path resulting in 0 connectivity.

The strength of the connectivity among all the nodes in a path must take into account the vulnerability of path elements (links and nodes) to attack or failure. We can account for this using the connectivity score described above by examining its value as we remove each node, link, or both, one at a time from a given path (which we assume here is the shortest path and has r_l nodes). For simplicity, we consider only the loss of nodes along the path.[7] We define the value ${}^{j}k_l$ as the connectivity value for information element a_l with the *j*th path node removed. We create a depletion vector, $\mathbf{L}_l$, whose elements are measures of how much connectivity is lost by the removal of each node, or $\mathbf{L}_l = [l_{l1}, l_{l2}, \cdots, l_{lr_l}]^t$, where $l_{lj} = k_l - {}^{j}k_l$ and r_l (as already noted) is the number of nodes in the shortest path that delivers information element a_l from any source to its destination.

The vector $\mathbf{L}_l$ represents the vulnerability of the shortest path and as such expresses the degree of uncertainty associated with retrieving information element a_l from network sources. The next step is to reduce the vector $\mathbf{L}_l$ to a scalar that can be used to reduce k_l, that is, to reflect the path uncertainty in terms of its connectivity value. A logical choice is the vector norm defined as

$$\|\mathbf{L}_l\| = \sqrt{\mathbf{L}_l^T \mathbf{L}_l}\,.$$

The vector norm measures the magnitude of the vector and therefore in this sense measures the magnitude of the potential depletions based on the shortest path. A large norm indicates that the depletion associated with removal of nodes from the shortest path is considerable. This means the connectivity associated with the shortest path is tenuous and should therefore be reduced accordingly. Conversely, if the norm is small, it reflects the fact that the shortest path is fairly robust and the reduction in the connectivity score should be minimal. Taking this into consideration, the adjusted connectivity for information element a_l from network sources to a single destination is calculated to be

[7] This approach, however, is equally valid if applied to links or both nodes and links.

$$k_l^* = k_l \left(1 - \frac{\|\mathrm{L}_l\|}{|\mathrm{L}_l|}\right)^{\frac{1}{\rho}},$$

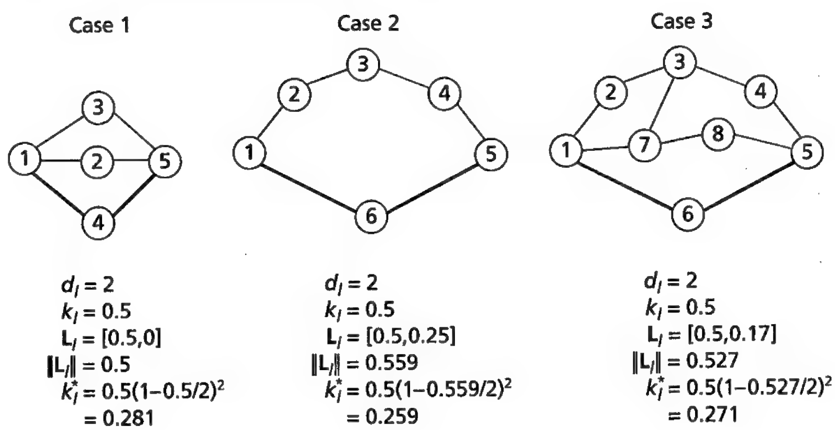
where $|\mathrm{L}_l|$ is the cardinality of the vector L_l and $0 < \rho \leq 1$ is the edge expansion parameter of the network that reflects the reliability of the network (Davidoff, Sarnak, and Valette, 2003). The edge expansion parameter is a generalisation of the clustering coefficient of a network, moving from considering single nodes to clusters of nodes. For example, Watts uses the clustering coefficient as part of the characterisation of small world networks (Watts, 1999).

The most reliable network is one in which every node is directly connected to every other node. Such a network is called 'complete' and leads to a value of $\rho = 1$. A value of ρ near to 1 thus implies that there are redundant paths in the network and, hence, high reliability. The edge expansion parameter ρ is calculated by considering clusters of nodes and how well they are connected to the rest of the network. Formally, for a finite network V, consider a subset U of V and let $|U|$ and $|V|$ represent the number of nodes in U and V, respectively. Let $E \subseteq V \times V$ be the edge set of V. For a given node v in V, define the neighbours of v as $\Gamma(v) = \{u \in V; (v,u) \in E\}$. For the cluster U, we can then define the neighbours of U as $\Gamma(U) = \cup_{v \in U} \Gamma(v)$. The boundary of the cluster U is defined as the neighbours of the cluster U less those nodes actually in the cluster U, i.e., $\partial U = \Gamma(U) - U$. Finally, the edge expansion parameter ρ is calculated by looking at the ratio of the size of the boundary of a cluster to the size of a cluster, considering all clusters within the network. In fact, we need only to consider clusters up to half the size of the total network to do this; thus,

$$\rho = \min \left\{ \frac{|\partial U|}{|U|} : U \subset V; 0 < \left| U \leq \frac{|V|}{2} \right| \right\}.$$

Figure 5.1 illustrates three simple cases, which are fragments of a larger network, for which the edge expansion parameter is $\rho = 0.5$.

Figure 5.1
Three Simple Connectivity Assessments



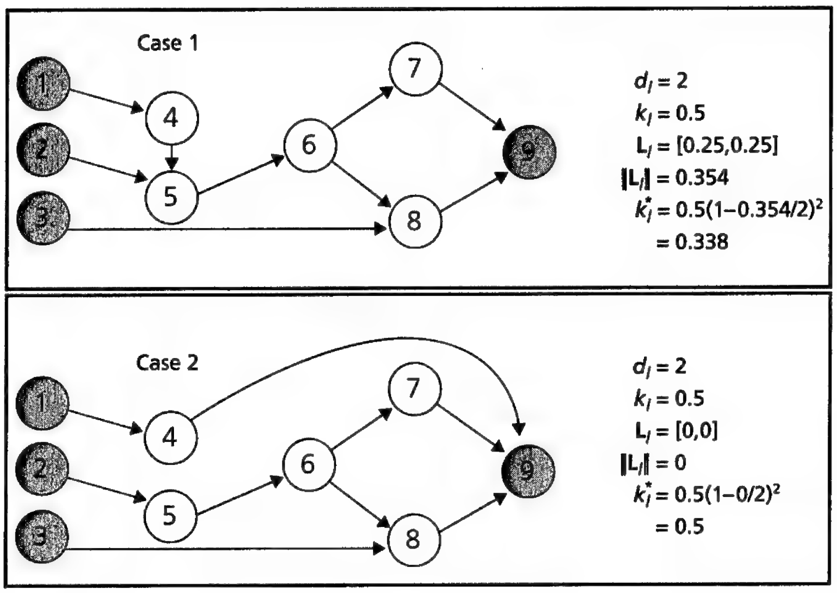
RAND *MG226-5.1*

We assume that the distance function, d_l, for information element a_l from a single source is measured as the number of nodes between the source and the destination. In addition, we set the decay factor as $\omega_l = 1$.

In all three cases, the initial connectivity score is 0.5. In case 1, removing the source (node 1) results in a total loss of connectivity reflected in the first entry in $\mathbf{L}_l$. Removing node 4 results in no loss of connectivity because there exists an alternative path, not including node 4, of the same length. This is reflected in the second entry in $\mathbf{L}_l$. In the second case, removing node 1 has the same effect as in case 1, but removing node 6 has the effect of reducing connectivity by 0.25. The entries in the vector $\mathbf{L}_l$ reflect results from the removal of both nodes in turn. In the last case, removing node 6 results only in a loss of 0.17 because of the existence of a shorter alternative path.

The examples in Figure 5.1 all have a single source for the information element a_l. A more realistic example would be one in which there are several sources for the same information element. Figure 5.2 examines two networks with three source nodes.

Figure 5.2
Connectivity Assessments with More Than One Source Node



RAND *MG226-5.2*

In case 1, the shortest path is from 3 to 8 to 9. If node 3 is eliminated, the shortest path has four links. The same thing happens if node 8 is removed, which results in a depletion vector that reflects a loss of half the connectivity score for both nodes. The effective connectivity drops from 0.5 to 0.338. In case 2, the addition of the link between nodes 4 and 9 provides an alternative path that is as short as the original path. This means that there is no loss in connectivity.

Accounting for the quality of information contained in accessibility, $X(k)$ entails replacing the binary count of the number of required information elements available in completeness with a connectivity score for each of the cluster critical information elements, or

$$X(k) = \begin{cases} \left(\dfrac{k}{C}\right)^{\xi} & C \neq 0 \\ 1 & \text{otherwise} \end{cases},$$

where $k = \sum_{l=1}^{C} k_l^*$ and C is, as before, the total number of information elements needed to render a decision within the cluster.[8]

Network Redundancy

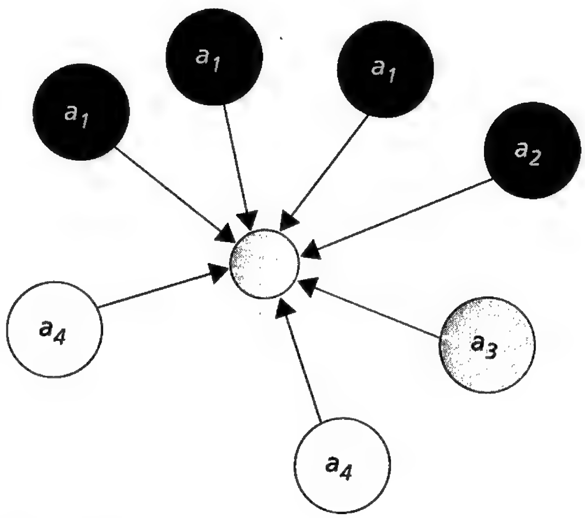
Network redundancy focuses on the reliability of the network—i.e., its ability to enable the delivery of information in the face of node loss, system outages, inefficient operating procedures, or some combination of all these. At the same time, a network can encourage the excessive delivery of information, thus causing delays as a result of the time and resources required to process it all. Consequently, network redundancy can be both a cost and a benefit of the network information flow.

In Figure 5.3, for our CEC cluster example, we assume that the node in the centre of the diagram is a decision node within the cluster, deciding an appropriate response to an incoming missile threat. The three nodes labelled a_1 provide position and velocity information; a_2 provides missile type information; and a_3 provides status information on friendly response systems (go, no-go). The nodes labelled a_4 and a_5 are also providing information; however, this information is not necessary to the node's decision to select a weapon system to engage the enemy missile.

The command nodes receive reports on the missile's position and speed from three sources. Because both will change over time, we can expect multiple reports from each. These multiple reports require combining in some way. We reflect the uncertainty associated with the position and speed of the missile by assuming they are random variables with known probability distributions, as discussed earlier. One method that allows for the sequential updating of probability

[8] Note that this formulation assumes that all information elements are equally important and that they are independent. We discuss dependent information elements in Chapter Three.

Figure 5.3
Node-Centric View of Information



RAND *MG226-5.3*

distributions is the one we have chosen: Bayesian updating.[9] Whatever method used, the degree to which the reports contribute to estimates close to ground truth and to narrowing the distribution variance can be considered a benefit in terms of redundancy.

However, all things being equal, the more sources of required information and the more frequent the reporting, the longer it takes for the decision node within the cluster to get a coherent view of the situation. This results from the fact that it takes time to process information that may or may not contribute to improving the quality of the estimates—essentially what is referred to as 'information overload'. In addition, some of the sources may provide disconfirming evidence. The value of the disconfirming evidence can be good or bad depending on the degree to which it reflects ground truth. Nevertheless, its presence increases uncertainty, requires time to evaluate, and

[9] In addition to Bayesian updating, the Dempster rule of combination and moving averages may be used to combine multiple observations. See Pearl (1987) and Shafer (1976).

therefore may decrease the quality of the estimates. Finally, it is also possible that raw data are processed before being sent, thus arriving at the command node as information that is time stamped with the time at which the processing ended. This possibility introduces an additional latency that contributes to uncertainty.

Unneeded Information

Dealing with information that is not needed is treated as a pure cost.[10] In Figure 5.3, the two information elements, a_4 and a_5, provide no useful purpose to the missile tracking and response mission. The costs of dealing with information of this type increases with the number of different information elements arriving at the command node and with their redundancy.

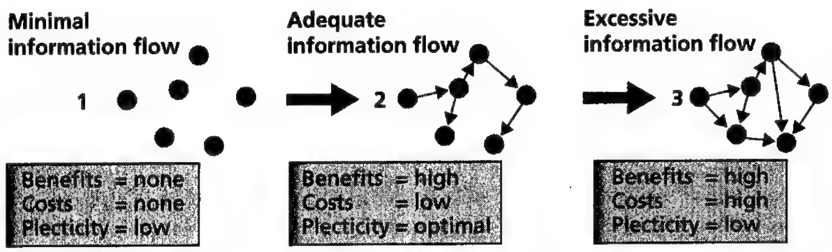
The Combined Effects

In the next section, we develop metrics for the measures just discussed. The result will be an overall metric for network plecticity. For networks with inadequate information flow, as with excessive information flow, we would expect low plecticity scores. The goal is to configure the information flow and clustering over a network with established link connectivity so as to maximise plecticity as measured in terms just discussed. If we assume a normalised plectic score, with 0 representing no plecticity and 1 representing maximum plecticity, then Figure 5.4 illustrates how the costs and benefits affect this score.

- **Minimal flow.** The first flow depiction in Figure 5.4 represents minimal information flow and a set of isolated nodes. Although depicted as having no information flow, in reality we would expect that there are a few sources of required information. However, there is no opportunity to share information, and we

[10] This is not always the case. In a rapidly evolving combat situation, information not needed at one moment can become crucial the next. In this case, it is important that the network be capable of adapting rapidly. However, there is still some cost associated with accepting and processing information that is not needed to prosecute the current operation.

Figure 5.4
Overall Network Plecticity



RAND *MG226-5.4*

assume that the decision nodes need not consult with each other before acting. The result is no benefits, no costs, and therefore a low plectic score.

- **Excessive flow.** Turning to the last flow depiction in Figure 5.4, the effects of information overload resulting from too much required and/or unneeded information results in low plectic values as well. The high benefits associated with a rich information flow are offset by the high costs of processing excessive information. Information can be shared directly among all the nodes.
- **Adequate flow.** Finally, the centre flow configuration in Figure 5.4 depicts reasonable redundancy of required information and limited unneeded information sources, thus resulting in optimal plectic values. The high benefits are associated with just the right amount of information flow and the costs associated with processing excessive information are therefore very low. The connectivity is rich, allowing for direct and indirect information sharing. The fewer channels per node result in fewer network ties and, therefore, a more manageable network.

The Benefits of Redundancy

As mentioned earlier, redundancy has both cost and benefit aspects, each requiring definition in metric form. Multiple reports of required information from several sources can increase the reliability of the estimates of information elements. At the same time, too many

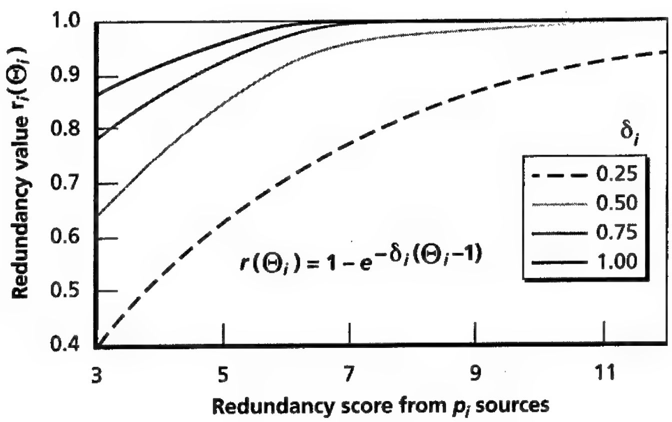
reports coming into, and being shared around, a cluster incur a cost because of information overload, reports of unneeded information, and possible disconfirming reports. We address the benefits first, but before beginning, we recognise the possibility that because the source of a rendered report is extremely reliable, its benefit might be considered equivalent to several reports from less reliable sources. This adds a complicating factor because the reliability of the sources of all reports must be assessed. Assuming the data are available to make this assessment, we can provide for this phenomena through suitable weighting.

First, we let $r_i(\Theta_i)$ be the benefit accruing from obtaining reports on the value of information element a_i from p_i sources, where $\Theta_i = \sum_{j=1}^{p_i} \theta_{i,j}$, and $\theta_{i,j} \in [1,\infty)$ measures the assessed reliability of the report on information element a_i from source $s_j (1 \le j \le p_i)$. This formulation ensures that $\Theta_i \ge 1$, as long as at least one report is received for information element a_i. Also, if all sources are minimally reliable, then $r_i(\Theta_i) = r_i(p_i)$, since $\theta_{i,j} = 1$ for all sources s_j. As with the accessibility metric X, we restrict $r_i(\Theta_i)$ to be between 0 and 1. In this case, $r_i(\Theta_i) = 0$ implies no benefit from redundancy. This result is equivalent to the case in which a reported estimate for information element a_i emanates from a single, marginally reliable source $(\theta_{i,j} = 1)$, or if no report is rendered, $r_i(\Theta_i) \to 1$ for some number of sources. A suitable model that reflects this behaviour is

$$r_i(\Theta_i) = \begin{cases} 1 - e^{-\delta_i(\Theta_i - 1)} & p_i \ge 1 \\ 0 & \text{otherwise} \end{cases}.$$

The parameter δ_i reflects the relative importance of the information element a_i. If a single report from an extremely reliable source arrives, it can be given a large weight so that $\Theta_i = \theta_{i,1}$ is large and $r_i(\Theta_i) \to 1$ for a single report. This metric therefore not only measures the effects of redundancy but also reflects the adequacy of the source of the report. Figure 5.5 illustrates how the value of the constant θ_i influences how rapidly redundancy and adequacy scores contribute to convergence.

Figure 5.5
The Effect of δ_i and Θ_i on the Benefits of Redundancy



p_i	$\theta_{i,1}$	$\theta_{i,2}$	$\theta_{i,3}$	Θ_i
3	1	1	1	3
2	2	4	–	6
1	9	–	–	9
2	2	10	–	12

RAND *MG226-5.5*

The table below the figure records the data used to construct the graphs. Note that for the third entry, only a single report source ($p_3 = 1$) exists, but it is considered more reliable than the two and three sources for entries 1 and 2. However, regardless of the redundancy scores, the impact of the information element importance scores is dramatic.

Having determined a redundancy benefit for each information element in a cluster's information set, we now combine the scores to arrive at an aggregate score for the required information set available across the cluster. Recall that the total number of required information elements across the whole network is N; the number critical to a cluster is C, where $C \leq N$; and the number of required information elements available within the cluster is n, where $n \leq C$. If we let the

vector $\Theta = [\Theta_1, \Theta_2, \cdots, \Theta_C]^T$ represent the value of reports received from the $\mathbf{P} = [p_1, p_2, \cdots, p_C]^T$ sources, we can construct a suitable normalised aggregate metric, $R(\Theta)$, as follows:[11]

$$R(\Theta) = \frac{1}{n}\sum_{i=1}^{C} r_i(\Theta_i)$$
$$= 1 - \frac{1}{n}\sum_{i=1}^{C} \gamma_i e^{-\delta_i(\Theta_i - 1)},$$

where $\gamma_i = 1$ if $p_i \geq 1$ and 0 otherwise. No penalty is assessed for missing information. This is accounted for in the accessibility score discussed earlier. In the case in which $n = C = 0$, we must have that $\Theta_i = 0$ and therefore $R(\Theta) = 0$—i.e., there is no redundancy benefit, even though the accessibility score is $X(k) = 1$.

Combining the Benefits

The next step is to combine the beneficial effects of information access, X, and redundancy, R, into a single metric for the cluster. To do this, we choose a conditional model. The benefits of redundancy depend on the information elements received by the cluster, in addition to the number of sources for each. The conditioning, however, is quite weak. For example, it is possible for a cluster to have perfect information access and score 0 for redundancy. Conversely, a cluster with very limited access can have a rather large redundancy score. But it is impossible to obtain positive redundancy benefit unless there is at least one report on at least one information element. A simple ratio, $R(\Theta)/X(k)$, exposes the desired relationship. However, the ratio is only bounded between 0 and 1 when $R(\Theta) \leq X(k)$ and $X_t(k) \neq 0$. We can modify the ratio using parameters to avoid a zero denominator

[11] Implicit assumptions in this form of aggregation are that (1) the value attributed to the reports is linear and (2) there is no value associated with the interactions among the reports. We discuss the issue of multi-attribute aggregation in Chapter Three.

and ensure the combined metric is bounded on [0,1]. We begin by setting

$$B\left[R(\Theta)\,|\,X(k)\right]=c\frac{\kappa+R(\Theta)}{\beta-X(k)}+d.$$

In this formulation, the parameter $\beta>1$ ensures a nonzero denominator and the parameter $\kappa\geq 0$ with the two constants c and d are used to ensure the combined metric is bounded between 0 and 1. The parameters, β and κ, reflect the relative importance placed on redundancy and completeness. The desired boundary conditions are $B(0\,|\,0)=0$ and $B(1\,|\,1)=1$. That is, obtaining the maximum redundancy given maximum access produces a maximum combined score, whereas it is impossible to achieve any redundancy given no access to the critical information elements.[12] The first condition yields

$$B(0\,|\,0)=c\frac{\kappa}{\beta}+d=0 \text{ and } d=-c\frac{\kappa}{\beta};$$

hence, we get

$$B\left[R(\Theta)\,|\,X(k)\right]=c\frac{\kappa+R(\Theta)}{\beta-X(k)}-c\frac{\kappa}{\beta}.$$

The second boundary condition yields

$$B[1\,|\,1]=c\frac{\kappa+1}{\beta-1}-c\frac{\kappa}{\beta}=\frac{c(\beta+\kappa)}{\beta(\beta-1)}=1;$$

therefore,

[12] Two other 'edge' conditions might be considered as well: $B(1\,|\,0)$, and $B(0\,|\,1)$. The former is not possible because it is impossible to accrue any benefit from redundant reports if critical information is inaccessible. The latter is equally impossible because it suggests that no benefit from redundancy is possible even though critical information is totally accessible. If at least one source reports on each critical information element, then $R(\Theta_i)>0$.

$$c = \frac{\beta(\beta - 1)}{\beta + \kappa} \text{ and } d = -\frac{\kappa(\beta - 1)}{\beta + \kappa}.$$

This gets us the final relationship,

$$B[R(\Theta) \mid X(k)] = \frac{(\beta - 1)[\kappa X(k) + \beta R(\Theta)]}{(\beta + \kappa)[\beta - X(k)]},$$

which is bounded between 0 and 1 and exhibits the required dependency between accessibility and the benefits of redundancy. Substituting

$$X(k) = \left(\frac{k}{C}\right)^{\xi} \text{ and } R(\Theta) = 1 - \frac{1}{n}\Sigma_{i=1}^{C} \gamma_i e^{-\delta_i(\Theta_i - 1)}$$

yields

$$B[R(\Theta) \mid X(k)] = \frac{(\beta - 1)\left(\kappa\left(\frac{k}{C}\right)^{\xi} + \beta\left(1 - \frac{1}{n}\Sigma_{i=1}^{C} \gamma_i e^{-\delta_i(\Theta_i - 1)}\right)\right)}{(\beta + \kappa)\left(\beta - \left(\frac{k}{C}\right)^{\xi}\right)}.$$

The Costs of Information Within a Cluster

The contribution of costs to plecticity within a cluster arises from three sources: (1) information overload, (2) disconfirming evidence, and (3) incomplete information. The latter cost is included in the calculation of the benefits associated with information accessibility. Disconfirming evidence has been covered previously as well. It arises as an issue when reports for disparate sources and sensors must be

combined to create a common operating picture. In this section, we focus exclusively on the costs of information overload. As mentioned earlier, information overload arises from too many sources of needed information and any source of unneeded information, which are both functions of redundancy. We begin with the costs of unneeded information.

Costs of Unneeded Information

In this analysis, the supply of unneeded information places a burden on the node receiving it and sharing it around the cluster. It has an immediate negative impact in that it must be processed or, at a minimum, interferes with the receipt of needed information. However, as more of it is supplied, its marginal impact is reduced in the same way email spam is dealt with in a modern office environment. Thus, a good function to model this behaviour is the exponential

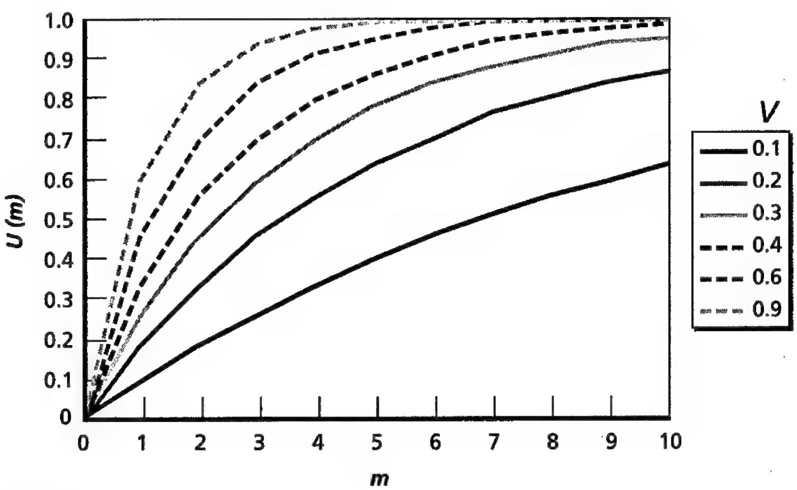
$$U(m) = 1 - e^{-\nu m},$$

where m is the number of sources of unneeded information and ν is a scaling parameter that reflects the rate at which unneeded information contributes to cost. This calculation then indicates the effect across the whole cluster, rather than at an individual affected node. In this case, no distinction is made between multiple sources of the same unneeded information and multiple sources of different information elements. Thus, the same cost results from the same information element supplied x times or x different information elements supplied once each. We show the influence of ν on the cost in Figure 5.6. As ν increases from zero, the saturation point is reached more rapidly.

Costs of Redundant but Needed Information

We now examine the effects of the cluster receiving too much needed information. As mentioned earlier, an overabundance of needed information contributes to information overload, increases the likeli-

Figure 5.6
Cost of Unneeded Information



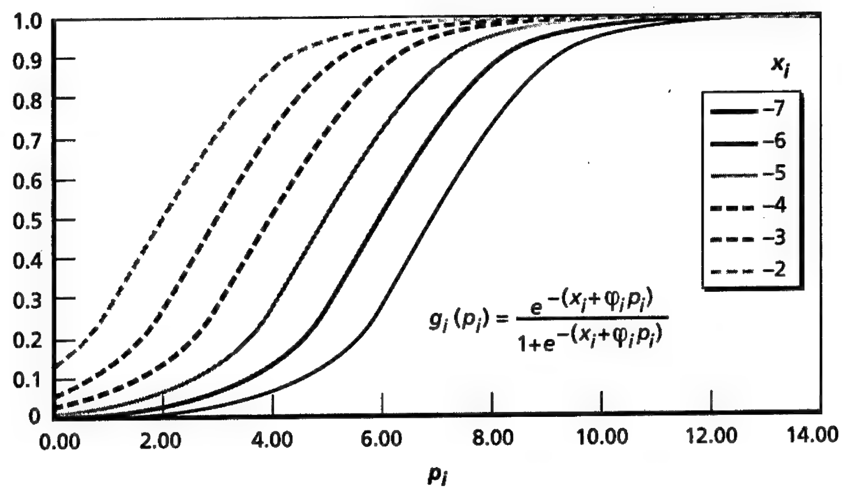
RAND *MG226-5.6*

hood that some of the information will be disconfirming, and therefore may cause delays in processing. The costs of information overload associated with needed information are generally minimal for low levels of redundancy. Indeed, at these levels, the benefits far outweigh the costs, as discussed earlier. However, at some point, costs rise sharply so that the marginal cost of an additional source of information is greater than the previous source. At some further point, this cost then levels off so that the marginal costs are minimal. This behaviour is best described using a logistics response function such as the following:

$$g_i(p_i) = \frac{e^{-(\chi_i + \varphi_i p_i)}}{1 + e^{-(\chi_i + \varphi_i p_i)}}.$$

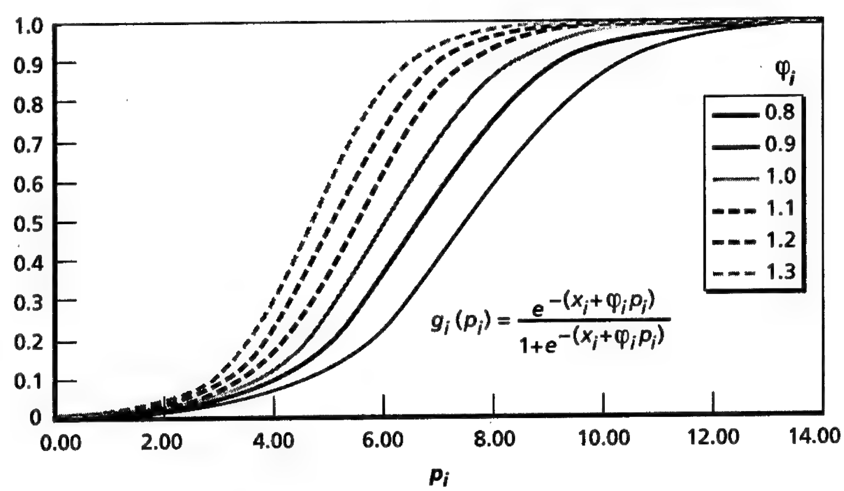
In this formulation, the p_i values are the number of sources for information element a_i as before and χ_i and φ_i are shaping parameters. We illustrate the influence of these parameters in Figures 5.7 and 5.8. The actual values will depend on the effects of receiving extra

Figure 5.7
The Costs of Redundancy for $\varphi_i = 1$



RAND *MG226-5.7*

Figure 5.8
The Costs of Redundancy for $\chi_i = -6$



RAND *MG226-5.8*

needed information. They can be assessed based on the point at which the extra sources of needed information begin to have a detrimental effect on operations at the node, the point at which the marginal cost of redundant information increases rapidly, and how soon after this the saturation point is reached—i.e., the point at which the marginal costs become negligible.

As was the case in calculating the overall benefit of redundancy, the costs of oversupply of each needed information type can be combined in a variety of ways. For simplicity, we expressed it here as a simple sum.[13]

$$G(\mathbf{P}) = \frac{1}{n}\sum_{i=1}^{C} g_i(p_i) = \frac{1}{n}\sum_{i=1}^{C} \gamma_i \frac{e^{-(\chi_i+\varphi_i+p_i)}}{1+e^{-(\chi_i+\varphi_i+p_i)}},$$

where $\mathbf{P} = [p_1, p_2, \cdots, p_C]^T$ and $\gamma_i = \begin{cases} 1 & \text{if } p_i \geq 1 \\ 0 & \text{otherwise} \end{cases}$.

Combining the Costs of Information for a Cluster

In considering the overall costs, a balance is struck between costs of needed and unneeded information. Unfortunately, the two are not independent. That is, the presence of one can greatly affect the cost of the other. For example, dealing with redundant needed information in the absence of any extraneous, noncontributing reports is clearly different than if the unneeded reports are present. However, the nature of the dependency is not clear. Consequently, we use a simple weighted linear sum of the two, or

$$\mathrm{O}[\mathrm{U}(m), \mathrm{G}(\mathbf{P})] = \alpha \mathrm{U}(m) + (1-\alpha)\mathrm{G}(\mathbf{P}),$$

where $0 \leq \alpha \leq 1$ is a relative weight parameter.

[13] The same assumptions made for the benefits of redundancy apply here as well.

Some may argue here that two functions have been used to model what is essentially the same cost: information overload. However, it is considered that these two types of information overload have different impacts on cluster effectiveness. Needed information affects the amount of information that needs to be processed, but there is also a greater chance of conflicting information, which places an additional burden on the cluster. Unneeded information is more easily dismissed, given that it is not essential to the user's needs.

Combining Costs and Benefits

The next step is to combine the costs and benefits of network plecticity for a cluster within the network, associated with the mission at hand. The term 'costs' suggests a simple cost-benefit analysis might be appropriate. In such a case, the benefit is divided by the cost, resulting in an assessment of the cost for a unit of benefit. However, in this analysis, we are not dealing with a true cost in the cost-benefit sense, but rather a cost more closely described as a penalty. We began this chapter by describing the characteristics of the network-plecticity metric, as illustrated in Figure 5.4. We assume each of the clusters in the network is logically connected to support a given mission. Plecticity for a cluster is then associated with the flow of information associated with that cluster. Both minimal (inadequate) flow and excessive flow should result in low plecticity, whereas 'optimal' (adequate) flow should result in high plecticity. Therefore, for each cluster W in the network, the measure of network plecticity $C(B,O)$ is calculated as follows:

$$
\begin{aligned}
C(B,O) &= B\left[R(\Theta) \mid X(k)\right]\left[1-O\left[U(m),G(\mathbf{P})\right]\right] \\
&= \frac{(\beta-1)\left[\beta R(?)+\kappa\, X(k)\right]\left[1-\alpha U(m)-(1-\alpha)G(\mathbf{P})\right]}{(\beta+\kappa)(\beta-X(k))}
\end{aligned}
$$

$$= \frac{(\beta-1)\left(\beta-\frac{\beta}{n}\sum_{i=1}^{C}\gamma_i e^{\delta_i(\Theta_i-1)}+\kappa\left(\frac{k}{C}\right)^{\xi}\right)\left[1-\alpha U(m)-(1-\alpha)G(\mathrm{P})\right]}{(\beta+\kappa)\left(\beta-\left(\frac{k}{C}\right)^{\xi}\right)}.$$

Overall Network Performance

The last step is to combine this redundancy-based plecticity with the benefits of collaboration to produce a metric that will assess the performance of networked decisionmaking headquarters. Collaboration measures the effects of information sharing across a cluster on information completeness and accuracy (i.e., bias and precision), whereas redundancy-based plecticity measures the effects of redundant information and the degree of information access. The former assesses the dynamic nature of the operation conducted on the network; the latter measures the effects of the underlying network structure and is therefore systemic. All the dependencies among the several components of collaboration and plecticity are not generally well understood. However, we know that high-quality performance requires good cluster knowledge and the means to share it and that scores in either category are penalised by deficiencies in the other. Therefore, the measure of total network performance is taken to be

$$\Omega(\Pi,\mathrm{K}_{Net})=\sum_{i=1}^{L}\left[C_i(B,O)K_{i,\kappa}\right]^{\omega_i},$$

where $\sum_{i=1}^{L}\omega_i=1$ and L is the total number of clusters across the network.

For values of $\Omega(\Pi,\mathrm{K}_{Net})$ close to 1.0, the network is performing well by producing the information required to take decisions within each of the clusters when required. However, this is not the whole story. The next step is to assess how well the combat mission is accomplished. As important as good decisions are, good combat out-

comes are the ultimate measure of the value of network-centric operations.

Summing Up

In assessing the effects of networking headquarters on increasing decisionmakers' knowledge and therefore improving decisions, we have analysed the network in terms of its static structure and of the dynamics associated with performing the operational mission. The former resulted in the development of a structural 'plecticity' metric for each cluster, and the latter in a dynamic 'knowledge' metric for each cluster. Both these metrics were developed by viewing a network of connected headquarters as a set of clusters within each of which all decision nodes (headquarters) share information. They are then combined to form a metric of overall network performance.

In the process of developing these metrics, we have appealed to information sciences, probability and statistics, estimation theory, complexity theory, combinatorics, and, of course, a large measure of heuristics. In the process, several terms were introduced as shaping parameters. For the most part, these parameters are designed to reflect the behaviour of both physical and cognitive phenomena. Where possible, we suggest methods for assessing reasonable values for these parameters. Nevertheless, we recognise that establishing methods for assessing these values is an open research question that will require considerable experimentation.

The aim of the work presented in this chapter is to contribute to the development of a theory of such complex information networks in order to stimulate both further theoretical development and experimentation. Although we include an application of the measures and metrics in Appendix C, there is still much more work to be done in progressing this new science.

CHAPTER SIX

Conclusion

At the outset, we argued that it is important that military planners responsibly test the emerging network-centric concepts before their adoption. Several observers concerned about the 'irrational exuberance' surrounding the claimed benefits of network-centric operations support this view as well.[1] They argue, as we do, that the claimed benefits may prove to be true but that analysts should strive to assist the military community in assessing them. This recommendation implies employing the full range of analytic techniques: models, simulations, exercises, and experiments. The problem, however, is the paucity of tools that will allow us to quantify the benefits of local collaboration and clustering across an information network. Although we make no claim that the methods reported here are definitive, they do represent an approach that draws on several disciplines to assess how well alternative operating procedures and network configurations contribute to the decisions made by headquarters that share information and thus develop shared awareness and collaboration.

The approach taken brings together two key ideas. The first idea comes from previous work by RAND that shows how Shannon entropy can be used as the basis of a quantified measure of the knowledge resident within a cluster of decisionmakers who share information. Such an approach allows the concept of full shared awareness to be precisely defined in terms of such clusters and also

[1] See, for example, Giffen and Reid (2003) and Barnett (1999).

permits the measure of benefit to be lifted from the information domain to the cognitive domain in terms of our process model of information age warfare. The second idea comes from Dstl research on the representation of command and control (and the other associated elements of C4ISR) in aggregate constructive simulation models of conflict. This concept has resulted in the Rapid Planning Process, which gives a basis in terms of mathematical algorithms for the representation of expert decisionmaking in fast-paced, fluid circumstances. These ideas are brought together using the idea of 'plecticity' drawn from our view of the network as a complex system. Combining collaboration and plecticity results in a total measure of the benefits and costs associated with a particular local collaboration and clustering across such a network. The measure captures the ability of the clusters to support the decisionmaking process at a key decision point, in terms of determining to what extent the distributed decisionmakers are within their 'comfort zones' in relation to the values of their key decision elements, which are shared across the clusters.

We have adopted an approach that first deals with the conceptually simplest case, when the information elements forming the basis of the decisionmaking in a cluster of the network all have the same distribution of uncertainty (hence, we assume they are all normally distributed). In this case, with full shared awareness across the cluster, the knowledge available to the cluster can be quantified on the basis of the variance of the key decision elements and their covariance, which builds up over time. This first part of the work highlights in particular the benefit to local knowledge of such covariance (i.e., the degree to which one element of information relates to another) in quantifying such knowledge. Such a measure thus relates closely to 'commonsense' ideas of knowledge in terms of understanding how a number of elements relate one to another.

We then deal with the more general case of when the information elements shared across a cluster have more general distributions of uncertainty. A number of approaches to this case are examined based on a mixture of empirical and theoretical ideas. By combining these ideas, it is possible to form a complete chain of quantification—from an initial network architecture and local collaboration,

through the formation of clusters across the network, through to the overall plecticity and performance of such a network. In this way, different possibilities for collaboration (and hence different future headquarters structures based on such distributed clustering and local decisionmaking) can be compared in terms of their total network performance. This comparison measures the ability of such distributed decisionmakers to make better decisions, based on better understanding of the critical information elements shared across collaborating clusters in the network.

APPENDIX A

The Rapid Planning Process

Gary Klein's Recognition Primed Decision (RPD) model emphasises situation awareness (SA) (Klein, 1989). The goal of the SA process is to provide the decisionmaker (the command agent) with an understanding of what is happening in the outside world. In particular, the command agent, through SA, tries to answer the question: 'Is the situation that I perceive in the outside world one that I recognise? Because if I do recognise the situation, then my experience (long-term memory) tells me immediately which course of action (CoA) I should adopt, given this situation.'

The focus of the SA process is thus on pattern matching—analysing the information available about the outside world and trying to match the perceived state of the world to one of an existing array of *patterns* held in the command agent's long-term memory. Each pattern is a representation of a *situation*, and each situation is linked directly to a CoA appropriate to that situation. This linkage, held in the command agent's long-term memory, represents the command agent's *experience* and is what enables the command agent to make decisions rapidly without recourse to extensive option generation and evaluation.

We model this behaviour through the Rapid Planning Process. The model thus comprises four main stages: (1) observation analysis and parameter estimation, (2) situation assessment, (3) pattern matching and preferred posture selection, and (4) posture transition. We discuss the first three in the context of a simple land operation

example. The headquarters model is concerned only with the overall process up to the decision.

Stage 1: Observation Analysis and Parameter Estimation

Stage 1 involves analysing the command agent's current observations of the battlespace, which comprise data received by the command agent via its sensors. The analysis of these data consists of data smoothing and parameter (mean and covariance) estimation. Where the variables are normally distributed, the data analysis is performed by a collection of dynamic linear models (DLMs). A DLM is a mathematical structure for short-term forecasting, modelling, and analysis of time-series processes with normal errors. (DLMs are fully described in West and Harrison, 1997.)

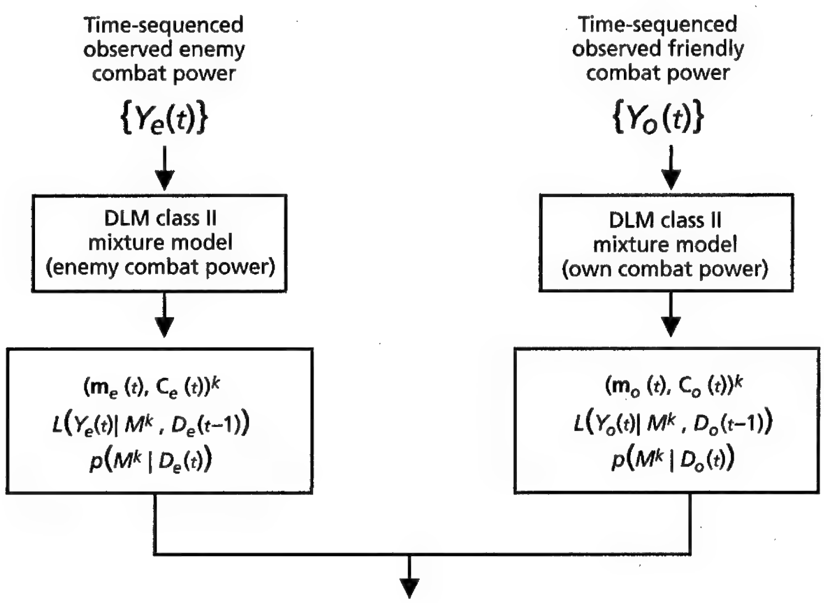
A Simple Land Operations Example

Figure A.1 illustrates the details of stage 1 of the Rapid Planning Process for this example. We assume decisionmaking is based on the perceived combat power ratio (PCPR) (see stage 3). The command agent deduces the PCPR from observations (via sensors) of two quantities in the local area of interest,[1] namely enemy combat power and friendly-force combat power. These two data input streams are analysed independently within the command agent via a pair of DLM class II mixture models—one model tracks the enemy combat power values while the other model independently tracks friendly-force combat power values.[2] In general, each class II mixture model comprises four separate DLMs: a 'standard' DLM, an outlier-generating DLM, a level change DLM, and a slope change DLM.

[1] The command agent's local area of interest is a circular region centred on the agent. The radius of this region is user specified. The agent's recognised picture covers only this region and is thus 'mobile'—that is, it moves with the agent.

[2] See West and Harrison (1997), §12.3.

Figure A.1
Stage 1: Observation Analysis and Parameter Estimation



RAND *MG226-A.1*

In each case, we look at a time series of observations of force levels (as assessed by a set of sensors or fed to the commander by an information source). Each DLM represents a predisposition by the commander to look at the series of estimates in a particular way, taking account of other contextual knowledge that may be available to him.

- **The 'standard' DLM** represents the assumption by the commander that nothing much is changing; he expects that the time series of observations will carry on at about the same level.
- **The outlier DLM** makes the assumption that the current observation is a significant deviation from the observations seen so far (causing a much higher variance in the series) but that the series is expected to settle back to the previous level.

- **The level change and slope change DLMs** represent an assumption by the commander that there will be significant change in level or slope (rate of change) of the series. For example, if the commander has access to his superior commander's plan, he may know that, at a certain time, additional friendly-force elements will move into his area of interest. He will thus be predisposed to look out for this when tracking the value of his own force strength over time.

Each of these DLMs is equivalent to a corresponding hypothesis by the commander about what is happening in his local area of interest while tracking the critical information element of force level over time: no change; a blip, which can be ignored; a step change; or a change in slope (growth or decay). When we have a vector of critical information elements making up the commander's conceptual space (also called the common relevant operating picture, or CROP), these hypotheses relate to the likely behaviour of the values of the critical information elements that form a vector characterising the conceptual space.

The 'standard' DLM is a first-order polynomial DLM, representing a system model M^1 that describes a constant level time series. The parameters estimated by the DLM are the mean and variance of the time series level denoted, at time t, by $m(t)$ and $C(t)$, respectively.

The other three DLMs in the mixture model are all second-order polynomial DLMs. The outlier-generating DLM represents a system model, M^2, that describes a transient in the time series. The level change DLM represents a system model, M^3, that describes a step change in the time series. The slope change DLM represents a system model, M^4, that describes a slope change in the time series. The parameters estimated by each of these three DLMs are the mean values of the level and the growth rate of the time series denoted, at time t, by vector $\mathbf{m}(t)$, and the associated covariance of the level and growth rates denoted, at time t, by matrix $\mathbf{C}(t)$.

The Common Relevant Operating Picture

The CROP (the local conceptual space) is spanned by the set of critical information elements. For our simple example, these relate to the local force ratio. In each case, the DLM formulation updates the assessment of where the commander perceives he is located within the space described by the vector of information elements. This corresponds to a multivariate normal distribution. The commander's fixed patterns correspond to particular 'areas' within this space that he considers important, such as good own force level and poor perceived enemy force level locally. To each of these fixed patterns is associated a particular mission, such as 'advance', representing the direct link between situation assessment and choice of feasible CoA required by the RPD approach. The overlap between the output from the DLM and the fixed patterns is used to update the probability that each of these patterns is the most relevant.[3]

In more detail, and taking as an example enemy and own force strengths as the factors forming the recognised picture, each DLM mixture model operates on an input time series, i.e., a sequence of observations received from external sensors.[4] For one mixture model, the input time series comprises observations of the enemy combat power in the command agent's local area of interest; this series is denoted by $Y_e(t)$ in Figure A.1 and comprises the sequence

$$\{Y_e(0), Y_e(1), \cdots, Y_e(t-1), Y_e(t)\}.$$

For the other mixture model, the input time series comprises observations of the friendly-force combat power in the command agent's local area of interest; this series is denoted by $Y_o(t)$ and comprises the sequence

$$\{Y_o(0), Y_o(1), \cdots, Y_o(t-1), Y_o(t)\}.$$

[3] An example of how this can be implemented is shown in Chapter 2 of Moffat (2002; also see p. 38).

[4] In this example, we focus on only a single critical information element: combat power.

Note that the observations in the two time series need not necessarily coincide because they are independent input streams.

Each DLM mixture model processes its associated time series of observations in the same way (and independently from the other DLM mixture models). We describe this process below for the enemy combat power time series; an analogous process operates in parallel for the friendly-force combat power time series. Figure A.1 shows the state of the parameter estimation process after the observations up to, and including, $Y_e(t-1)$ have been processed by the DLM mixture model and before the next observation, $Y_e(t)$, is processed. To process the next enemy combat power observation, $Y_e(t)$ is fed into the DLM mixture model and analysed. The DLM algorithms follow the Bayesian methods developed in West and Harrison (1997). At each stage of the process, a probability is computed for each of the commander's hypotheses (corresponding to one of the DLMs). These probabilities are tracked over time to assess whether we are approaching the boundary of the 'OK' state, i.e., the probability of no change has declined significantly. The following are key outputs of the mixture model:

- **Updated estimates of the system model parameters.** These estimates now take into account the new observation $Y_e(t)$. There are four sets of these estimates, denoted $(\mathbf{m}_e(t), \mathbf{C}_e(t))^k$, where $k \in [1,4]$ is the DLM type. One set of estimates is produced by each DLM in the mixture model. The particular values $(\mathbf{m}_e(t), \mathbf{C}_e(t))^j$ are the current estimates of the mean and covariances of the enemy combat power (level and growth) on the assumption that system model M^j represents the time series seen to date.
- **Likelihood estimates for each system model.** This is the likelihood that the observation $Y_e(t)$ would have been obtained from each system model. There are four of these, one for each DLM in the mixture model. The likelihoods are denoted $L(Y_e(t) \mid M^k, D_e(t-1))$, where $D_e(t-1)$ represents all observations seen up to, but not including, the current observation, $Y_e(t)$. This is repeated for the friendly-force estimates.

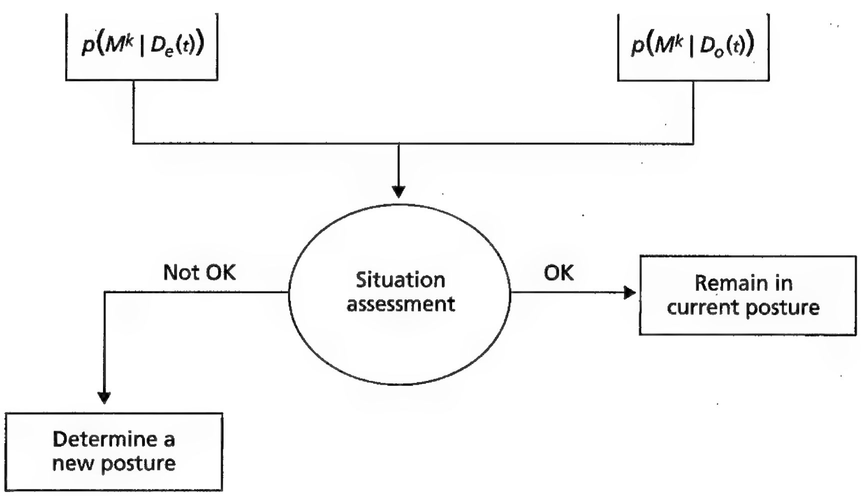
- **Posterior probabilities that each hypothesis is correct.** The posterior probabilities $p(M^k \mid D_e(t))$ are the probability that model M^k best describes the time series of observations seen up to time t. This is repeated for the friendly-force estimates.

The posterior probabilities $p(M^k \mid D_e(t))$ (for the enemy combat power observations) and $p(M^k \mid D_o(t))$ (for the friendly force combat power observations) are updated on a continuous basis as part of the command agent's sensing cycle.

Stage 2: Situation Assessment

The means, covariances, likelihood estimates, and posterior probabilities are input to stage 2 in the Rapid Planning Process. Figure A.2 illustrates the processes in this stage. At each command and control cycle (which runs independently of the sensing cycle), the command

Figure A.2
Stage 2: Situation Assessment



RAND *MG226-A.2*

agent performs an SA to decide whether the perceived situation, based on the sensor observations made to date, is currently 'OK' or 'Not OK'. The situation assessment is performed in two steps.

Step 1—OK/Not OK Assessment

The first step of the SA considers the enemy combat power and friendly-force combat power observations separately, as follows. Examining each DLM mixture model:

- If the 'standard' DLM has the highest posterior probability, the situation is deemed OK. This conclusion is based on the fact that the combat power observed is currently showing a steady level.[5]
- If any of the other three DLMs (the outlier, the level change, or the growth change models) has the highest posterior probability, the situation is deemed Not OK. This conclusion is based on the fact that the combat power observed has changed from a steady level.

Step 2—Initial Situation Assessment

Step 1 generates an OK/Not OK result from each DLM mixture model. In the second step, we combine these results, using Table A.1, to determine an overall assessment of the current situation. This corresponds to the 'storytelling' level of SA discussed by Klein (1989).

[5] In West and Harrison's version of the DLM class II mixture model (West and Harrison, 1997, §12.3), the 'standard' model is the linear growth model (the second-order polynomial DLM). It should now be clear why, in our case, we actually need the standard model to be the constant model (the first-order polynomial DLM), representing a system model that describes a constant level time series. It is because a linear growth model used as the standard model (the OK model) might remain the most likely model throughout—so that we would interpret the situation as remaining OK—while actually tracking a steady drift of combat power values across a wide range—so that the situation therefore might not always be OK from a PCPR perspective. The only OK situation is the one in which the combat power observations are remaining more or less constant—hence the use of the constant (first-order) DLM.

Table A.1
Initial Situation Assessment Matrix

Friendly-Force Combat Power Mixture Model	Enemy Combat Power Mixture Model: OK	Enemy Combat Power Mixture Model: Not OK
OK	OK	Not OK
Not OK	Not OK	Not OK

Thus, the overall SA is OK only if the situation is OK with respect to both the enemy and friendly-force combat power observations. In each of the other cases, one or another, and possibly both, of the SAs are Not OK because there has been a significant change in the enemy and/or friendly-force combat power and the overall SA is deemed Not OK.

The idea behind the SA described here is to provide an initial OK/Not OK alert to the command agent. If the situation is OK, the command agent carries on doing whatever it is currently doing—it remains in its current posture; there is no need to do any (stage 3) pattern matching and preferred posture selection, because everything is currently OK.

If, however, the situation is Not OK, then only in this case does the command agent need to go into stage 3 of the Rapid Planning Process and do some pattern matching to find out if a change in posture is required.

If the situation is Not OK, the command agent invokes stage 3 of the Rapid Planning Process model. Some key data items[6] are passed to stage 3—namely $m_e(t)$ and $m_o(t)$, the current best estimates of the enemy and own force combat power values, respectively, and their associated variances, $C_e(t)$ and $C_o(t)$. These 'best' estimates are

[6] In this version of the Rapid Planning Process model, only the means and variances of the combat power values are used in stage 3. We do not forward to stage 3 any of the additional information that is actually available at the end of stage 2, namely the growth rate and its variance (in the case of second-order polynomial DLMs) and knowledge of which system models are the better descriptors of each combat power time series. Future enhancements to the model will likely make use of this additional information.

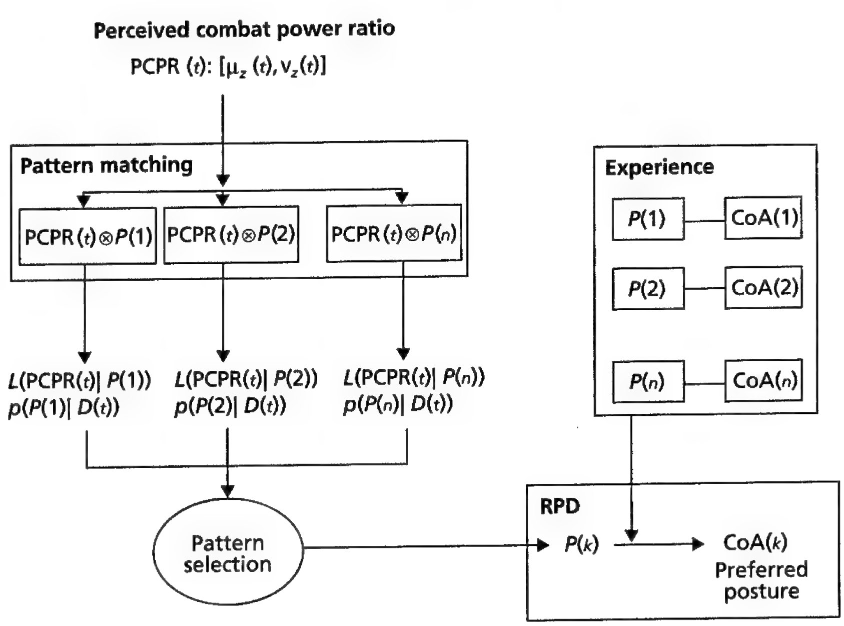
the values output by the DLM, in each mixture model, which currently has the highest posterior probability.

Stage 3: Pattern Matching and Course of Action Selection

Stage 3 of the Rapid Planning Process model attempts to recognise the extant battlespace situation, based on the data received by the command agent, and identify the posture (CoA) appropriate to this situation. Figure A.3 illustrates the process.

As mentioned earlier, the inputs to stage 3 are the current best estimates of the enemy and own force combat power values, respectively, and their associated variances. From these, we calculate the

Figure A.3
Stage 3: Pattern Matching and Course of Action Selection



RAND *MG226-A.3*

PCPR at the current time t, denoted PCPR(t). The PCPR(t) is a random variable with probability density having a time-dependent mean, $\mu_z(t)$, and a standard deviation, $\nu_z(t)$. We depict these elements in Figure A.3. The PCPR(t) distribution characterised by its mean and standard deviation is input to the main pattern-matching process.

The pattern-matching process (denoted by symbol ⊗ in Figure A.3) compares the PCPR(t) distribution against a number of patterns, denoted $P(k)$. Each pattern is a representation of one possible situation that could exist in the battlespace. Comparing is aimed at selecting the most likely pattern given the PCPR(t) being compared. The comparison (pattern match) of PCPR(t) against a given pattern, $P(k)$, yields two outputs:

- $L(\mathrm{PCPR}(t) \,|\, P(k))$: The likelihood that PCPR(t) would have been obtained had the situation in the battlespace been the one represented by pattern $P(k)$.
- $p(P(k) \,|\, D(t))$: The posterior probability that pattern $P(k)$ is the one that best represents the situation in the battlespace, given the time series of (enemy and own force combat power) observations seen up to time t (i.e., the current situation).

Having calculated the posterior probability of each pattern $P(1), P(2), \cdots, P(n)$, we select the pattern $P(k)$ with the highest posterior probability as the one that best represents the situation extant in the battlespace. The situation has now been 'recognised'.

The next step—and the essence of the RPD model of decision-making—is to invoke the decisionmaker's experience and map the recognised situation to an appropriate CoA. In Figure A.3, experience is represented by the set of one-to-one mappings between patterns $P(i)$ and CoA(i) for all $i \in [1,n]$ stored in the command agent's long-term memory. Thus, the selected pattern $P(k)$, representing the recognised situation, leads directly to the selection of an appropriate CoA, namely CoA(k).

CoA(k), selected in this way, is referred to as the *preferred posture*. It is the posture that the command agent's experience says is most appropriate, given the situation recognised in the agent's local

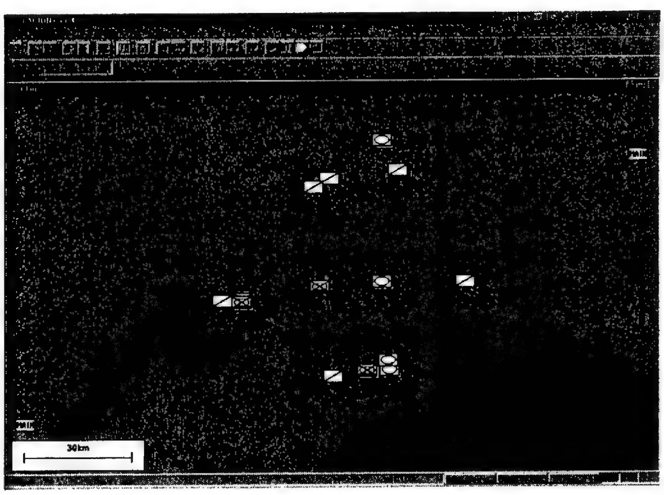
area of interest. The preferred posture is then passed into stage 4 (posture transition) of the Rapid Planning Process model. We do not make use of stage 4 of the Rapid Planning Process in the method proposed in this report. Our processing terminates with the selection of CoA(k).

Moffat (2002) details the mathematical development of these algorithms for the general case of a conceptual space with several factors.

Application

The following is an example application. The modelling test bed used is CLARION+, an experimental test bed developed by the Defence Science and Technology Laboratory (Dstl) to examine the effect of such decisionmaking on combat behaviour. Figure A.4 is a screen

Figure A.4
CLARION+ Screen Image of Land-Air Interaction



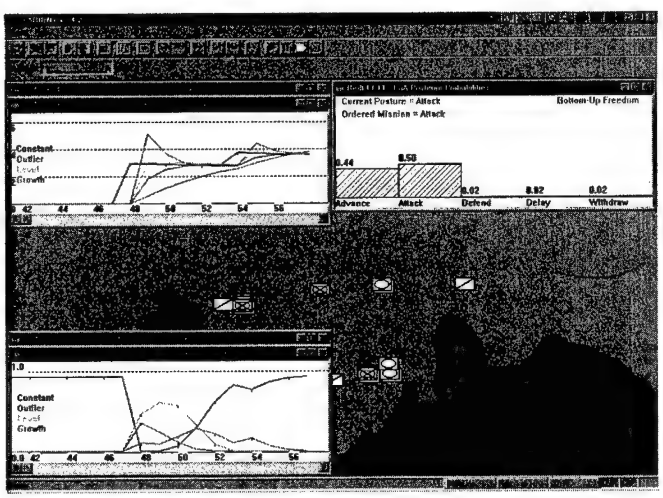
RAND *MG226-A.4*

image from CLARION+ that shows a campaign-level land-air interaction between two forces (Red and Blue) in which Red, using a bold command strategy developed by a genetic algorithm, fixes Blue in the south and then flanks north to exploit a hole in Blue's defence. The boxes with a single diagonal line marking are airborne sensors that help to generate the operational picture and assessment of enemy intent on which the plan is based.

For one of the brigades in the circle, the dynamics of the Rapid Planning DLMs used to assess the level of enemy force strength in the local area of interest of the brigade are shown in Figure A.5.

At the top left-hand part of the figure, the mean values of enemy force strength assessed in the local area are shown for each of the four mixture models (standard, outlier, level change, slope change).[7] These values vary with time along the x-axis and grow as the brigade encounters an enemy group in its local picture.

Figure A.5
Rapid Planning Type II Mixture Model Depiction



RAND *MG226-A.5*

[7] In Figure A.5, these models are called constant, outlier, level, and growth, respectively.

The bottom left-hand corner of the figure displays the time-varying Bayesian probabilities that each of the four models is a correct assessment of the situation. These probabilities also vary with time along the x-axis. The most likely hypothesis moves from the standard model, through the outlier model, to a realisation that there is a level change in enemy combat power occurring in the local area. The constant model later supersedes this again.

At the top right is a display of the probability that each of the fixed patterns (and hence the associated CoA) is the best pattern match for the current perceived situation, for that brigade, at the time the simulation test bed was paused. The possible courses of action are advance, attack, defend, delay, or withdraw. The figure shows that at the time the simulation was paused, the local commander favoured advancing or attacking.

APPENDIX B

Information Entropy

Claude Shannon first introduced information entropy in 1948. In the early 1940s, it was generally believed that noise limited the flow of information through a channel. That is, if one decreased the probability of error in the received message, the true rate of data transmission decreased. Consequently, an error-free message could only occur if transmission ceased! Shannon disproved this theory. He was able to show that, in fact, if a channel had nonzero *capacity* (calculated from the noise of the channel), an arbitrarily low probability of error could be achieved as long as the transmission rate was below the channel capacity. He also argued that random processes such as speech and music had an irreducible complexity below which signal compression was impossible. He referred to this as *entropy* and further claimed that if the entropy at the source of a communication channel was less than its capacity, an arbitrarily low error rate could be achieved.[1]

It is this reference to communications as a stochastic or random process that leads to its application in the field of statistics. In his book on information theory, Solomon Kullback (1978, p. 1) cites several sources to support his argument that the statistical theory of communications is synonymous with communications theory and that communications theory and information theory are also synonymous. Because probability distributions describe the uncertainty associated with mutually exclusive and collectively exhaustive events,

[1] This summary draws on Cover and Thomas (1991), Blahut (1987), and Kullback (1978).

it is natural to ask about uncertainty's complement, that is, what is *known*, or the amount of information available. This leads us to the modern use of information theory as a measure of the average information available in a probability distribution.

A Statistical Theory of Information

Suppose $X = \{x_1, x_2, \cdots, x_n\}$ is a discrete random variable with probability mass function $P(X = x_i) = p_i$. Each of the x_i represents an event (as do conjunctions and disjunctions of the x_i), the occurrence of which imparts information. What we seek is a measure for the *amount* of information imparted. It seems logical to assume that this amount, whatever it is, is inversely proportional to the likelihood that the event will occur, or

$$I(x_i) \approx \frac{1}{p_i}.$$

For example, the fact that the sun rose this morning imparts no information, because we knew it all along. That is, the likelihood of its occurrence is 1.0. Conversely, being told that you have just won the lottery conveys considerable information because it is an unlikely event.

In a 1928 paper, Ralph Hartley was the first to suggest the use of the logarithm in a measure of information by defining a measure of information to be the logarithm of the number of possible symbol sequences (Hartley, 1928). Shannon picked up the idea of using the logarithm as the proportionality constant and suggested that the amount of information in the occurrence of an event is

$$I(x_i) = \log\left(\frac{1}{p_i}\right) = -\log(p_i).$$

This was a particularly good choice because it is closely related to the concept of data compression, as we shall see next. Shannon was

concerned with the output of a discrete information source where each of the x_i represents a source output that occurs with probability p_i. For this reason, the base 2 logarithm was used and information was measured in terms of *bits*.[2] However, in this work and elsewhere, we use the base e and measure information in terms of 'natural units', or *nats*.

The next step is to assess the *average* or *expected* information in the probability distribution. This quantity is referred to as information entropy or Shannon entropy and is calculated as

$$E\left[-\log\left(P(X)\right)\right] = H(X) = -\Sigma_{i=1}^{n} p_i \log p_i .$$

The quantity $H(X)$ represents the mean information content in $P(X)$ or the amount of uncertainty in $P(X)$. The latter interpretation implies that information entropy is a function of the variance of a distribution. This is the case and is readily evident in continuous distributions. If the base 2 logarithm is used, it is also the number of bits required, on average, that are used to describe the random variable, X.

It is interesting to note that for discrete random variables, entropy is indeed bounded. A lower bound (maximum certainty) occurs when $p_i = 1$ and $p_j = 0$ for all $j \neq i$. In this case,

$$H(X) = -1\log 1 - (n-1)0\log 0 = 0.$$[3]

Therefore, the average information is 0 nats when there is no uncertainty. This is consistent with the earlier definition of information. At the other end of the spectrum, complete uncertainty exists when all events are equally probable. The entropy calculation in this case is

---

[2] It turns out that one bit of information is the minimum information required to resolve the uncertainty in a situation with two equally probable alternatives.

[3] It can be shown that

$$\lim_{x \to 0} x \log x = 0 .$$

$$H(X) = -\sum_{i=1}^{n} \frac{1}{n} \log \frac{1}{n} = \frac{1}{n} \sum_{i=1}^{n} \log n = \log n .$$

Consequently, for discrete random variables, the average information of the probability mass function is bounded, or $H(X) \in [0, \log n]$. We show later that this is never true for continuous random variables; that is, the entropy for continuous random variables is unbounded.

Differential Entropy

The foregoing discussion assumed that the random variable was discrete, and we were able to show that the entropy of its probability mass function was bounded. In information theory, this is the equivalent to stating that the information source is discrete and that it generates discrete information at a finite rate. In contrast, the entropy of the density function for a continuous random variable is unbounded. In information theory, this is equivalent to a continuous information source that can assume any one of an uncountable infinite number of amplitude values, thus requiring an infinite number of binary bits for its complete specification. Because this is never possible, its entropy is unbounded.

Suppose now that X is a continuous random variable with probability density function $f(x)$. The differential entropy of X in nats is defined to be

$$H(X) = -\int_{-\infty}^{\infty} f(x) \log f(x) dx .$$

Unlike the entropy of a discrete random variable, the entropy of a continuous random variable is unbounded. We can illustrate this fact by approximating the continuous probability density, $f(x)$, with a probability mass function, $p(x)$, that is constant on intervals of width Δx. The approximating probability density function has probability p_j on the jth interval. To ensure that

$$\sum_j p_j = 1,$$

we set $p_j = p(x_j)\Delta x$, where x_j is a point in the jth interval such that $p(x_j)\Delta x$ is the area under $p(x)$ in the jth interval. The entropy of the approximating probability distribution is

$$\begin{aligned} H(p) &= -\sum_j p_j \log p_j \\ &= -\sum_j p(x_j)\Delta x \log\left[p(x_j)\Delta x\right] \\ &= -\sum_j p(x_j)\Delta x \log\left[p(x_j)\right] - \sum_j p(x_j)\Delta x \log[\Delta x] \\ &= -\sum_j p(x_j)\Delta x \log\left[p(x_j)\right] - \log[\Delta x]. \end{aligned}$$

Now, if we let $\Delta x \to 0$, the summation converges to an integral, but $\log[\Delta x] \to -\infty$. Because there is no way to avoid this divergence, the entropy of a continuous random variable is always unbounded.

Differential entropy can also be negative. For example, consider a random variable, X, distributed uniformly from 0 to a. Its density function is

$$f(x) = \begin{cases} 1/a & \text{if } 0 \le x \le a \\ 0 & \text{otherwise} \end{cases}.$$

The differential entropy therefore is

$$H(X) = -\int_0^a \frac{1}{a} \log \frac{1}{a} dx = \log a.$$

Note that for $a < 1$, $H(X) = \log a < 0$. $H(X)$ is also unbounded at $a = 0$.

Suppose X is a continuous random variable distributed exponentially with mean $1/\lambda$. The density function for X therefore is

$$f(x) = \lambda e^{-\lambda x}, \ x \ge 0.$$

The differential entropy is

$$\begin{aligned} H(x) &= -\int_0^\infty \lambda e^{-\lambda x} \log \lambda e^{-\lambda x} dx \\ &= 1 - \log(\lambda_j) \\ &= \log\left(\frac{e}{\lambda_j}\right) \end{aligned}$$

The differential entropy for several probability distributions have been tabulated by Thomas Cover and Joy Thomas and can be found in their book, *Elements of Information Theory* (1991).

APPENDIX C

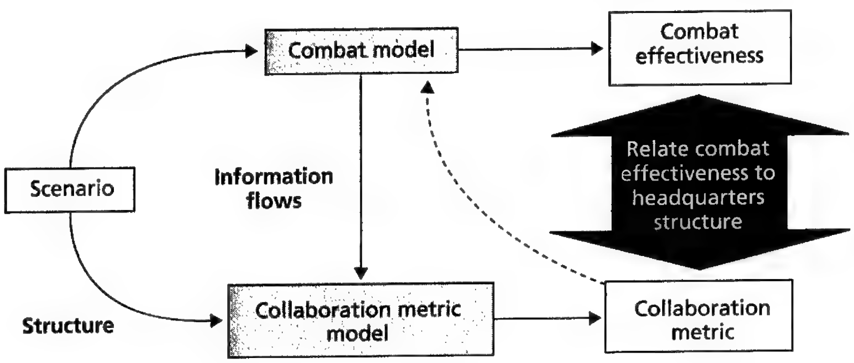
Application to a Logistics Network[1]

This appendix records the application of both the plecticity and collaboration metrics, with some extensions, to the logistics example discussed in the main text. As mentioned earlier, it is important to assess the effects of improved decisionmaking on combat outcomes. The measures and metrics we have developed are designed to assess the degree to which sharing information among headquarters in a clustered network contributes to improved decisionmaking. The ultimate measure of this effect is how well the friendly forces achieved their mission, i.e., combat effectiveness. Consequently, Dstl has developed a spreadsheet version of the information-sharing model, the Collaboration Metric Model (CMM), which is used to calculate both the plecticity and collaboration metrics described in the text for specific clusterings of decisionmaking nodes across an information network.

Figure C.1 summarises a methodology for assessing alternative command and control processes, using a combination of combat simulation modelling and the CMM. Information flows recorded in the simulation model are used as inputs for the CMM. The CMM results may then be used to select preferred network structures as inputs to the simulation model, as depicted in Figure C.1 by the dashed line. It is then possible to relate Measures of Command and Control

[1] The analysis presented in this appendix is primarily the work of RAND colleague Chris Pernin while on secondment to Dstl.

Figure C.1
Assessing the Effects of Information Sharing on Combat Effectiveness



RAND *MG226-C.1*

Effectiveness of the network clustering and Measures of Force Effectiveness, thus illustrating the relationships between information sharing and combat effectiveness.

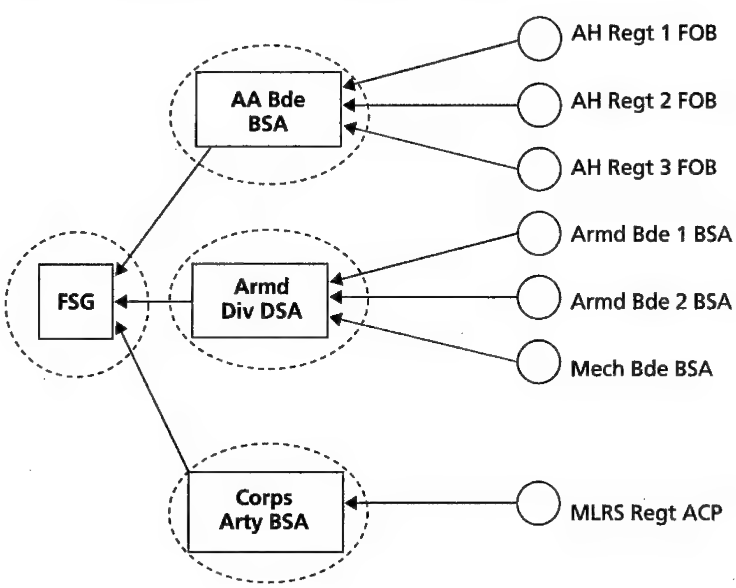
The CMM can handle up to 10 decision nodes, 10 information elements, and 10 information sources (see Figure 2.2 in Chapter Two for illustrations of these different network elements). This capacity allows a reasonable representation of a rather robust headquarters. The metrics discussed in the text form the basis of the Overall Network Performance metric calculated by the model and include both the static systemic measures of plecticity and the dynamic measures of collaboration. These are combined to arrive at a single metric to assess the effects of collaboration and plecticity across a cluster of information-sharing entities.

Cases Examined

Three logistics command and control structures were assessed using the CMM. The decision made in all cases is the logistics allocation decision described in Chapter Two; except that in this application, the resupply of ammunition, not fuel, was the focus. The first case is a supply-driven network similar to the 'push' sustainment model depicted in Figure 2.3 in Chapter Two. In this case, denoted S, the

Forward Support Group (FSG); the Air Assault Brigade, Brigade Supply Area (AA Bde BSA); the Armoured Division, Divisional Supply Area (Armd Div DSA); and the Corps Artillery, Brigade Supply Area (Corps Arty BSA) all form decision nodes, as shown by the rectangles in Figure C.2. However, there is no information sharing to form a common perception; thus, each of these decision nodes is a degenerate 'cluster' consisting of one node, shown by the dashed ellipses. Information on logistics demand is sent to these second and third line units from the Attack Helicopter Regiment Forward Operating Base (AH Regt FOB); the Armoured Brigade, Brigade Supply Area (Armd Bde BSA); the Mechanised Brigade, Brigade Supply Area (Mech Bde BSA); and the Multiple-Launch Rocket System Regiment Ammunition Control Point (MLRS Regt ACP). These information sources are shown as circles in Figure C.2. The amount supplied is based on a set expectation of use.

Figure C.2
A Supply-Driven Information Network: Case S

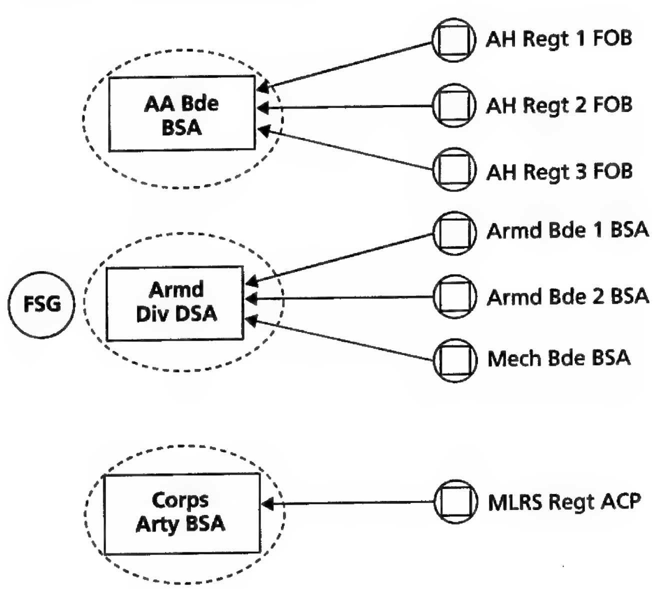


RAND *MG226-C.2*

The next two cases are demand driven and denoted as D1 and D2. Demand driven means that the units anticipate their supply requirements and decide how much resupply to request, or 'pull', from their arbiters at the next command echelon. How well they do this depends on their ability to share information, as we will see.

In the first demand case, D1, depicted in Figure C.3, each first and second line unit (10 units in total) sends its demand for an asset, which is met by the resource manager. The managers deal with each demand separately (i.e., they do not cross-correlate demands from different subordinate units). In this case, there are 10 decision nodes, each of which forms an isolated cluster of size 1.

Figure C.3
A Demand-Driven Information Network with No Information Sharing: Case D1

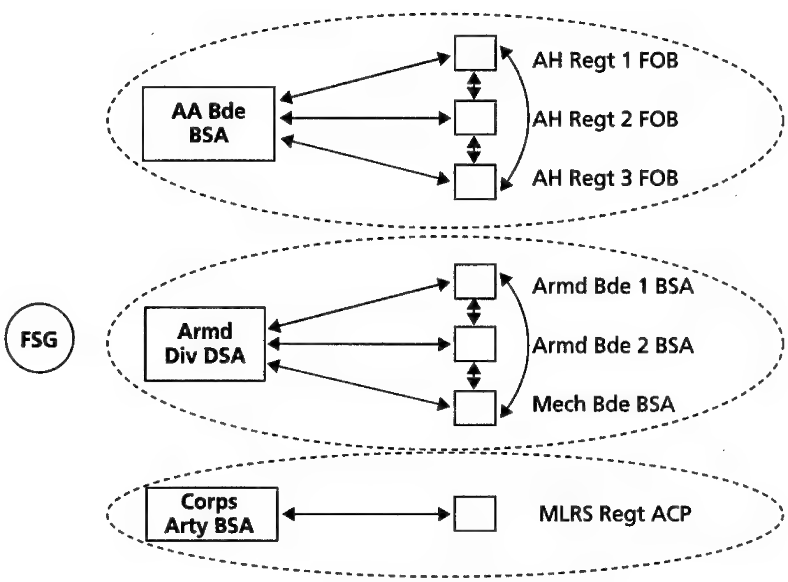


RAND *MG226-C.3*

The second demand-driven network, case D2, is depicted in Figure C.4. In this final case, each of the three second line logistics units is clustered with its subordinates into a full information sharing and shared awareness cluster. The superior units use their knowledge of all their subordinates' information elements to update their perception of the current status and needs of each unit.

The first two cases, S and D1, are extremes in logistic decision-making. The first case uses doctrine to push materiel to the units, regardless of unfolding events. The amount being pushed to the units is decided a priori and is not updated over time. The second case uses a daily update of what was consumed to resupply stocks to previous

Figure C.4
A Demand-Driven Information Network with Information Sharing: Case D2



RAND *MG226-C.4*

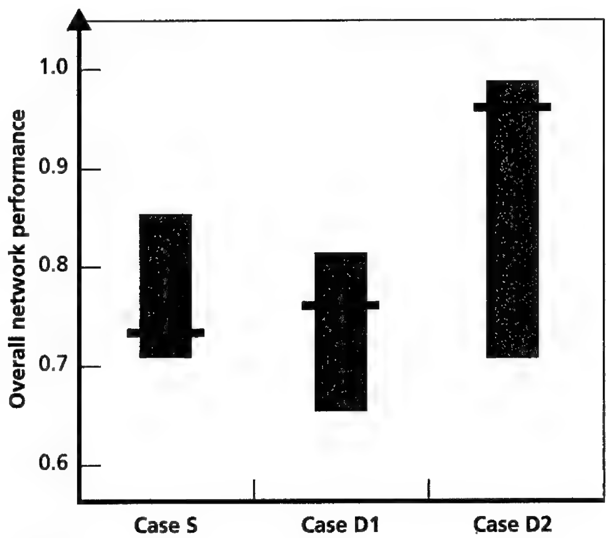
levels. The third case (D2) is a variant on the second case but contains additional clustering of information. This case uses three clusters that contain the 10 decision nodes.

Discussion and Results

Figures C.5 and C.6 show two metrics calculated by the CMM. Figure C.5 is the Overall Network Performance (combining collaboration and plecticity) for each of the three options. These values can range from 0 (very poor performance) to 1 (excellent performance). The shaded region defines the minimum and maximum of the value over the 24-hour scenario; the black bar shows the average over time. From Figure C.5, we can see that the most significant difference arises from the clustering of decision nodes. Cases D1 and D2 have the same information elements and number of decision nodes. They differ crucially in the number of clusters sharing information. In the former case, each logistics unit is introduced to one information element and develops an understanding of the logistics consumption based on that information. In the latter, the decision nodes are able to access information from neighbouring units that help build a better understanding of the situation. Even though both demand cases seem to have a much better understanding of the information elements over time compared with the supply-driven case, it is only when the information is shared among decision nodes that the increase in Overall Network Performance becomes evident. In this example, the sharing of information provides a greater increase to the overall ability of the network to perform compared with the location of the decisionmaking.

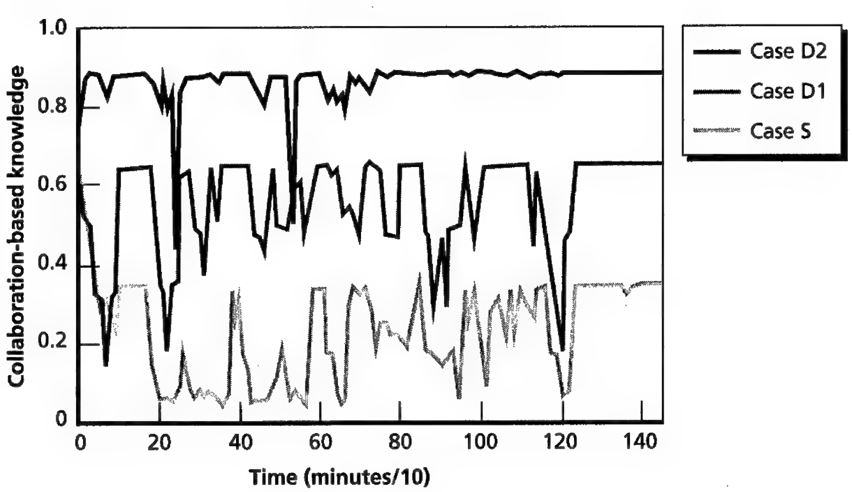
Figure C.6 records the knowledge derived from collaboration only, that is, the dynamic elements of the information network. The collaboration-based knowledge metric measures the knowledge gained from the dynamics of the information network, as discussed in Chapter Four.

Figure C.5
Overall Network Knowledge



RAND *MG226-C.5*

Figure C.6
Collaboration-Based Knowledge



RAND *MG226-C.6*

There are two main differences among the three cases. The first is the variation within each data set. A comparison of the three cases reveals that the upper case (case D2) has much less variation between adjacent points than the lower two cases (cases D1 and S). The enhanced clustering in case D2 compared with D1 has perhaps relieved the uncertainty of unexpected changes in the information elements. A reduced sensitivity to changes in the information elements is reflected in a less volatile and smoother line. The knowledge of three units engaged in a sudden change in their supply level will be more *understandable* or *palatable* to a commander than if only one unit experiences that change.

The second difference among the data is the level of collaboration-based knowledge. Case S exhibits the lowest knowledge level, reflecting the large differences between the average doctrinal use compared with the actual use during combat. The two demand cases provide enhanced knowledge compared with the supply case because the baseline is much more closely related to the actual use. The difference between the two demand cases provides the value of shared information between peers. The information elements and baselines are the same in both demand cases. However, the system values calculated through the dynamic linear models (see Appendix A) are much closer, and hence have enhanced knowledge, in the case of the more collaborative network. In this example, the three-cluster demand-driven network (case D2) provides the clearest picture of the consumption of the subordinate units.

Bibliography

Albert, R., and A-L. Barabási, 'Statistical Mechanics of Complex Networks', *Reviews of Modern Physics*, Vol. 74, January 2002.

Alberts, D. S., et al., *Network Centric Warfare*, 2nd edition, US Department of Defense, Command and Control Research Program, 2002.

Alberts, D. S., J. J. Gartska, R. E. Hayes, and D. A. Signori, *Understanding Information Age Warfare*, US Department of Defense, Command and Control Research Program, 2001.

Anderson, P. W., 'More Is Different', *Science*, Vol. 177, 1972, pp. 393–396.

Ayyub, B. M., and R. H. McCuen, *Probability, Statistics, & Reliability for Engineers*, New York: CRC Press, 1997.

Barnett, T. P. M., 'The Seven Deadly Sins of Network-Centric Warfare', *Proceedings* [US Naval Institute], January 1999.

Blahut, R. E., *Principles and Practice of Information Theory*, Boston: Addison-Wesley, 1987.

Bridgman, P. W., *Dimensional Analysis*, New Haven, Conn.: Yale University Press, 1922.

Burington, R. S., and D. C. May, *Handbook of Probability and Statistics with Tables*, Sandusky, Ohio: Handbook Publishers, 1958.

Butte, A. J., and I. S. Kohane, 'Mutual Information Relevance Networks: Functional Genomic Clustering Using Pairwise Entropy Measurements', Children's Hospital Informatics Program and Division of Endocrinology, Boston, 1999.

'The Cooperative Engagement Capability', *Johns Hopkins Technical Digest*, Vol. 16, No. 4, 1995.

Cover, T. M., and J. A. Thomas, *Elements of Information Theory*, New York: Wiley, 1991.

Davidoff, G., P. Sarnak, and A. Valette, *Elementary Number Theory, Group Theory, and Ramanujan Graphs*, Cambridge, UK: Cambridge University Press, 2003.

de Neufville, R., *Applied Systems Analysis: Engineering Planning and Technology Management*, New York: McGraw-Hill, 1990.

Endsley, M. R., 'Toward a Theory of Situational Awareness in Dynamic Systems', *Human Factors*, Vol. 37, No. 1, 1995, pp. 32–64.

Feltham, S., C. Sheppard, and C. Cooper Chapman, 'The Effect of Picture Quality on Individual Situational Awareness and Mission Effectiveness', paper presented at the International Symposium on Military Operational Research, United Kingdom, August 2003.

Fishburn, P. C., *Additive Utilities with Incomplete Product Set: Applications to Priorities and Assignments*, Baltimore, Md.: Operations Research Society of America, 1967.

Gell-Mann, M., 'What Is Complexity?' *Complexity*, Vol. 1, No. 1, 1995.

_____, 'Let's Call It Plectics', *Complexity*, Vol. 1, No. 5, 1995/1996.

Giffen, R. E., and D. J. Reid, 'A Woven Web of Guesses, Canto One: Network-Centric Warfare and the Myth of the New Economy', in *Proceedings 8th International Command and Control Research and Technology Symposium*, Washington, D.C., 17–19 June 2003.

Hartley, R. V. L., 'Transmission of Information', *Bell System Tech Journal*, Vol. 7, July 1928, pp. 535–563.

Jackson, B. W., and D. Thoro, *Applied Combinatorics with Problem Solving*, Reading, Mass.: Addison-Wesley, 1989.

Klein, G., 'Recognition Primed Decisions', in W. B. Rouse, ed., *Advances in Man-Machine Systems Research*, Vol. 5, Greenwich, Conn.: JAI Press, 1989, pp. 47–92.

Kolmogorov, A. N., 'Three Approaches to the Quantitative Definition of Information', *Problems of Information Transmission*, Vol. 1, 1965, pp. 4–7.

Kreyszig, E., *Introductory Functional Analysis with Applications*, New York: Wiley, 1978.

Kullback, S., *Information Theory and Statistics*, Gloucester, Mass.: Peter Smith, 1978.

Moffat, J., *Command and Control in the Information Age. Representing Its Impact*, London: The Stationery Office, 2002.

_____, *Complexity Theory and Network Centric Warfare*, US Department of Defense, Command and Control Research Program, 2003.

Pearl, J., *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, San Francisco: Morgan Kaufmann, 1987.

Pecht, M., ed., *Product Reliability, Maintainability, and Supportability Handbook*, New York: CRC Press, 1995.

Perry, W., R. W. Button, J. Bracken, T. Sullivan, and J. Mitchell, *Measures of Effectiveness for the Information-Age Navy: The Effects of Network-Centric Operations on Combat Outcomes*, Santa Monica, Calif.: RAND Corporation, MR-1449-Navy, 2002.

Perry, W., D. Signori, and J. Boon, *Exploring Information Superiority: A Methodology for Measuring the Quality of Information and Its Impact on Shared Awareness*, Santa Monica, Calif.: RAND Corporation, MR-1467-OSD, 2004.

Shafer, G., *A Mathematical Theory of Evidence*, Princeton, N.J.: Princeton University Press, 1976.

Shannon, C. E., 'A Mathematical Theory of Communications', *Bell Systems Tech Journal*, Vol. 27, 1948, pp. 379–423, 623–656.

Solé, R. V., and B. Luque, 'Statistical Measures for Complexity for Strongly Interacting Systems', *Physical Review E*, Vol. 1, 27 August 1999.

Sporns, O., and G. Tononi, 'Classes of Network Connectivity and Dynamics', *Complexity*, Vol. 7, No. 1, 2002, pp. 28–38.

Watts, D. J., *Small Worlds: The Dynamics of Networks Between Order and Randomness* (Princeton Studies in Complexity), Princeton, N.J.: Princeton University Press, 1999.

West, M., and J. Harrison, *Bayesian Forecasting and Dynamic Models* (Springer Series in Statistics), 2nd edition, Berlin: Springer-Verlag, 1997.

Wolpert, D., and W. Macready, 'Self-Dissimilarity: An Empirical Measure of Complexity', working paper, Santa Fe Institute, Santa Fe, N.M., 1997.

Yoon, K. P., and C-L. Hwang, 'Multi Attribute Decision Making: An Introduction', Sage University Papers Series, No. 07-104, Sage Publications, London, 1995.